Introduction to Cosmology (SS 2014 block course)

Lecturers: Markus Pössel & Simon Glover

1 Formation of structure: linear regime

- Up to this point in our discussion, we have assumed that the Universe is perfectly homogeneous on all scales. However, if this were truly the case, then we would not be here in this lecture theatre.
- We know that in reality, the Universe is highly inhomogeneous on small scales, with a considerable fraction of the matter content locked up in galaxies that have mean densities much higher than the mean cosmological matter density. We only recover homogeneity when we look at the distribution of these galaxies on very large scales.
- The extreme smoothness of the CMB tells us that the Universe must have been very close to homogeneous during the recombination epoch, and that all of the large-scale structure that we see must have formed between then and now.
- In this section and the next, we will review the theory of structure formation in an expanding Universe. We start by considering the evolution of small perturbations that can be treated using linear perturbation theory, before going on to look at which happens once these perturbations become large and linear theory breaks down.

1.1 Perturbation equations

- We start with the equations of continuity

$$\frac{\partial \rho}{\partial t} + \vec{\nabla} \cdot (\rho \vec{v}) = 0, \tag{1}$$

momentum conservation (i.e. Euler's equation)

$$\frac{\partial \vec{v}}{\partial t} + \left(\vec{v} \cdot \vec{\nabla}\right) \vec{v} = -\frac{\vec{\nabla} p}{\rho} + \vec{\nabla}\Phi \tag{2}$$

and Poisson's equation for the gravitational potential Φ:

$$\nabla^2 \Phi = 4\pi G \rho. \tag{3}$$

- We next split up the density and velocity in their homogeneous background values ρ_0 and $\vec{v}_0$ and small perturbations $\delta\rho$, $\delta\vec{v}$. If we let $\vec{r}$ represent physics coordinates, then our unperturbed velocity is simply

$$\vec{v}_0 = H\vec{r}, \tag{4}$$

i.e. it is the Hubble flow.

- To first order in our small perturbations, the continuity equation becomes

$$\frac{\partial\left(\rho_0+\delta\rho\right)}{\partial t}+\vec{\nabla}\cdot\left(\rho_0\vec{v}_0+\delta\rho\vec{v}_0+\rho_0\delta\vec{v}\right)=0. \tag{5}$$

This can be simplified by noting that the unperturbed density and velocity must also satisfy a continuity equation

$$\frac{\partial\rho_0}{\partial t}+\vec{\nabla}\cdot\left(\rho_0\vec{v}_0\right)=0. \tag{6}$$

Hence, our perturbation equation becomes

$$\frac{\partial\delta\rho}{\partial t}+\vec{v}_0\cdot\vec{\nabla}\delta\rho+\rho_0\vec{\nabla}\cdot\delta\vec{v}+\delta\rho\vec{\nabla}\cdot\vec{v}_0=0. \tag{7}$$

(Note that the $\nabla\rho_0$ term vanishes due to the homogeneity that we have assumed for our unperturbed state).

- If we define the density contrast

$$\delta\equiv\frac{\delta\rho}{\rho_0}, \tag{8}$$

then we can write this in a more compact form as

$$\dot{\delta}+\vec{v}_0\cdot\vec{\nabla}\delta+\vec{\nabla}\cdot\delta\vec{v}=0. \tag{9}$$

- From the momentum conservation equation, we obtain the relationship

$$\frac{\partial\delta\vec{v}}{\partial t}+\left(\delta\vec{v}\cdot\vec{\nabla}\right)\vec{v}_0+\left(\vec{v}_0\cdot\vec{\nabla}\right)\delta\vec{v}=-\frac{\vec{\nabla}\delta p}{\rho_0}+\vec{\nabla}\delta\Phi, \tag{10}$$

which can be simplified to

$$\frac{\partial\delta\vec{v}}{\partial t}+H\delta\vec{v}+\left(\vec{v}_0\cdot\vec{\nabla}\right)\delta\vec{v}=-\frac{\vec{\nabla}\delta p}{\rho_0}+\vec{\nabla}\delta\Phi, \tag{11}$$

- Finally, from the Poisson equation we have

$$\nabla^2\delta\Phi=4\pi G\rho_0\delta. \tag{12}$$

- We now introduce comoving coordinates $\vec{x}\equiv\vec{r}/a$, and comoving peculiar velocities, $\vec{u}\equiv\delta\vec{v}/a$. Our spatial derivative transforms as

$$\vec{\nabla}_r=\frac{1}{a}\vec{\nabla}_x. \tag{13}$$

Our time derivative, on the other hand, transforms as

$$\frac{\partial}{\partial t}+H\vec{x}\cdot\vec{\nabla}_x\rightarrow\frac{\partial}{\partial t}. \tag{14}$$

- In comoving coordinates, our perturbation equations become

$$\begin{aligned} \dot{\delta} + \vec{\nabla} \cdot \vec{u} &= 0, && (15) \\ \dot{\vec{u}} + 2H\vec{u} &= -\frac{\vec{\nabla}\delta p}{a^2\rho_0} + \frac{\vec{\nabla}\delta\Phi}{a^2}, && (16) \\ \nabla^2 \delta\Phi &= 4\pi G \rho_0 a^2 \delta, && (17) \end{aligned}$$

  where for simplicity we write $\vec{\nabla}_x$ simply as $\vec{\nabla}$.

- To close this set of equations, we also need an equation of state linking the pressure and density fluctuations:

$$\delta p = c_s^2 \delta\rho = c_s^2 \rho_0 \delta. \quad (18)$$

1.2 Density perturbations

- By combining our first two perturbation equations, we can derive the following second-order differential equation of the density contrast:

$$\ddot{\delta} + 2H\dot{\delta} = \left(4\pi G \rho_0 \delta + \frac{c_s^2 \nabla^2 \delta}{a^2}\right). \quad (19)$$

- To solve this, we start by decomposing δ into a set of plane waves:

$$\delta(\vec{x}, t) = \int \frac{\mathrm{d}^3 k}{(2\pi)^3} \hat{\delta}\left(\vec{k}, t\right) e^{-i\vec{k}\cdot\vec{x}}. \quad (20)$$

- Our Fourier amplitudes then obey the equation

$$\ddot{\hat{\delta}} + 2H\dot{\hat{\delta}} = \hat{\delta}\left(4\pi G \rho_0 - \frac{c_s^2 k^2}{a^2}\right). \quad (21)$$

- In the limit $k \rightarrow 0$ (i.e. the long wavelength limit), this reduces to

$$\ddot{\hat{\delta}} + 2H\dot{\hat{\delta}} = 4\pi G \rho_0 \hat{\delta}, \quad (22)$$

  which we recognise as the equation for a damped harmonic oscillator.

- In an $\Omega_m = 1$ Universe, we can write this equation as

$$\ddot{\hat{\delta}} + 2H\dot{\hat{\delta}} = \frac{3}{2} H^2 \hat{\delta}. \quad (23)$$

  (In an $\Omega_m \neq 1$ Universe, things are a little more complex, but at high redshift $\Omega_m \simeq 1$).

- We now try a solution of the form $\hat{\delta} \propto t^n$. This yields the equation

$$n(n-1)\frac{\hat{\delta}}{t^2} + 2Hn\frac{\hat{\delta}}{t} = \frac{3}{2}H^2\hat{\delta}. \tag{24}$$

For $\Omega_m = 1$, we know that $H(t) = 2/3t$, and so this equation becomes

$$n(n-1)\frac{\hat{\delta}}{t^2} + \frac{4}{3}n\frac{\hat{\delta}}{t^2} = \frac{2}{3}\frac{\hat{\delta}}{t^2}. \tag{25}$$

This equation has a non-trivial solution only when n satisfies

$$n^2 + \frac{n}{3} - \frac{2}{3} = 0. \tag{26}$$

This equation has solutions $n = 2/3$ and $n = -1$, corresponding to a growing mode $\hat{\delta} \propto t^{2/3}$ and a decaying mode $\hat{\delta} \propto t^{-1}$.

- It is convenient to express the evolution of δ with redshift in terms of the current value, δ_0, and a term known as the linear growth factor, $D_+(z)$:

$$\delta(z) = \delta_0 D_+(z). \tag{27}$$

For an Einstein-de Sitter Universe, $D_+(z) = (1+z)^{-1}$. For other cosmological models, we have the rather more complicated expression:

$$D_+(z) = \frac{1}{1+z}\frac{5\Omega_{\rm m}}{2}\int_0^1 \frac{{\rm d}a}{a^3 H(a)^3}. \tag{28}$$

- Observations of the CMB show us that at $z \sim 1000$, the perturbations in the gas component have amplitudes that are of order 10^{-5}. If these perturbations then grow as $\hat{\delta} \propto t^{2/3}$, then by the time we reach redshift zero, they will have grown by at most a factor of 1000, and will still be of order 1%.

- Clearly, perturbations in the gas component alone cannot account for the highly inhomogeneous density distribution we see around us. So how does this structure form?

- **Dark matter** provides a resolution to this conundrum. Perturbations in the dark matter couple to the radiation field only through their gravitational influence (rather than by direct scattering, as is the case for the gas), and hence can be much larger than the gas perturbations without overly perturbing the CMB. By starting with much larger perturbations, we can reach the $\delta \sim 1$ regime much sooner, allowing us to form the observed structures.

1.3 Jeans length, Jeans mass

- From Equation 21, we see that the source term for our density perturbation equation is positive only if

$$k > k_{\rm J} \equiv \frac{2\sqrt{\pi G \rho_0}}{c_{\rm s}}. \tag{29}$$

In other words, we will get growing perturbations only if they have wavenumbers that satisfy this criterion.

- An alternative way to express this criterion is in terms of a critical wavelength, defined as

$$\lambda_{\rm J} \equiv \frac{2\pi}{k_{\rm J}} = c_{\rm s}\sqrt{\frac{\pi}{G\rho_0}}. \tag{30}$$

  This critical value is known as the **Jeans length**. Only perturbations with wavelengths greater than the Jeans length will grow.

- Physically, we can understand the existence of this critical length scale by considering the balance between gravity and thermal pressure. If we take a small part of the pre-galactic gas and perturb it adiabatically, its density and temperature will increase. It will therefore be over-pressured relative to the surrounding gas, and the pressure gradients that we have created will try to smooth out the perturbation. Our perturbation will survive and grow only if its **self-gravity** – i.e. the gravitational force acting on the perturbation due to the perturbation's own mass – is larger than the pressure forces acting to smooth out the perturbation.

- It should be plain that for very small perturbations, with very low masses, pressure will overcome gravity. Similarly, it should be clear that on very large scales, gravity will win. There must therefore be some intermediate scale at which we go from being pressure-dominated to being gravity-dominated. This critical scale is just the Jeans length.

- We can also define a critical mass scale to go along with our critical length scale. This mass scale is known as the **Jeans mass** and is given by[1]

$$M_{\rm J} = \frac{4\pi}{3}\rho_0\left(\frac{\lambda_{\rm J}}{2}\right)^3. \tag{31}$$

- What happens if instead of gas, we consider dark matter? Most viable dark matter candidates are effectively collisionless, and hence have no sound speed *per se*. Does this mean that we can simply set $c_{\rm s} = 0$, and hence conclude that perturbations on all scales are unstable?

- For **cold dark matter** (CDM), this is actually a pretty good approximation. However, on very small scales it breaks down due to a phenomenon known as **free streaming**. This refers to the fact that our collisionless dark matter particles have a non-zero velocity dispersion. If their velocities are larger than the escape velocity of our perturbation, then they will simply stream away from the overdensity before it can undergo gravitational collapse.

[1] Note that there is a certain arbitrariness in our choosing to compute the mass within a sphere of radius $\lambda_{\rm J}/2$, and not, say, a sphere of radius $\lambda_{\rm J}$ or a cube of side length $\lambda_{\rm J}$. Consequently, the Jeans mass is a somewhat fuzzy concept, and should best be thought of as simply giving us a guide to the critical mass of an unstable perturbation. In practice, for perturbations with $M \sim M_{\rm J}$, we generally need numerical simulations in order to determine the ability of the perturbation to collapse and the timescale on which this occurs, particularly if the latter is comparable to the current expansion timescale.

- A careful analysis of this phenomenon leads one to derive an expression very similar to that for the Jeans length, only with the velocity dispersion of the dark matter in place of the sound speed. However, for CDM, the velocity dispersion is very small, and hence the Jeans mass and Jeans length of the dark matter are also very small; for instance, Diemand et al. (2005, Nature, 433, 389) show that for WIMP dark matter, the lowest mass dark matter halos should have masses of the order of an Earth mass.

1.4 Perturbations in a radiation-dominated Universe

- Up to this point, we have implicitly been assuming that the Universe is matter dominated. However, our initial density perturbations come into existence during the inflationary epoch and hence spend the first part of their life growing during the radiation-dominated era.

- In principle, correct treatment of perturbation growth during the radiation-dominated era requires a relativistic treatment of the governing equations. In practice, provided we are dealing with small perturbations, a non-relativistic treatment suffices.

- If we ignore pressure gradients (i.e. consider scales much larger than the Jeans length), then the governing equation for the growth of density perturbations in the radiation-dominated case can be derived in a similar fashion to that in the matter dominated case if we make the substitutions $\rho \rightarrow \rho + p/c^2$ in the continuity equation, and $\rho \rightarrow \rho + 3p/c^2$ in the Poisson equation. Using the fact that $p = \rho c^2/3$ for radiation, we find that

$$\ddot{\hat{\delta}} + 2H\dot{\hat{\delta}} = \frac{32\pi}{3} G\rho_0 \hat{\delta}. \tag{32}$$

- We can rewrite this equation in terms of the Hubble parameter as

$$\ddot{\hat{\delta}} + 2H\dot{\hat{\delta}} = 4H^2 \hat{\delta}, \tag{33}$$

  Since $H = 1/2t$ in the radiation-dominated era, we find that we again have two solutions: a growing mode with $\hat{\delta} \propto t$ and a decaying mode with $\hat{\delta} \propto t^{-1}$. (Note that in deriving these soluations, we have assumed that $\Omega = 1$. This is always a good approximation during the radiation-dominated era).

- In terms of the scale factor, our growth mode is $\hat{\delta} \propto a^2$; hence, long wavelength perturbations grow much faster with increasing a in the radiation-dominated era than in the matter-dominated era, where they evolve only as $\hat{\delta} \propto a$.

- This is for perturbations on scales large enough that pressure forces are irrelevant. What happens on smaller scales? In the case of perturbations in the radiation or in the baryons (which are strongly coupled to the radiation at this point), the behaviour is fairly clear. We define a Jeans length as before,

$$\lambda_{\rm J} = c_{\rm s} \sqrt{\frac{\pi}{G\rho_0}}, \tag{34}$$

but in this case our sound-speed $c_{\rm s} = c/\sqrt{3}$, where c is the speed of light, and so

$$\lambda_{\rm J} = c\sqrt{\frac{\pi}{3G\rho_0}}. \tag{35}$$

- If we compare this number to the Hubble radius, $r_{\rm H} = c/H$, we find that $\lambda_{\rm J}/r_{\rm H} = \sqrt{8\pi^2/9} \simeq 3$; in other words, perturbations on scales comparable to the size of the observable Universe are suppressed during the radiation-dominated era.

- What about the dark matter? This does not couple directly to the radiation, and hence does not feel the radiation pressure. However, the growth of perturbations on scales $\lambda \ll r_{\rm H}$ is nevertheless suppressed, for a reason that we will now explain.

- If we consider scales $r \ll \lambda_{\rm J}$, then we can ignore any perturbations in the radiation component and treat it simply as a flat background. In this limit, the equation describing the growth of perturbations in the dark matter then becomes

$$\ddot{\hat{\delta}} + 2H\dot{\hat{\delta}} = 4\pi G\rho_{\rm m}\hat{\delta}. \tag{36}$$

  Since we are in the radiation-dominated regime, we can write $H^2 = 8\pi G(\rho_{\rm m} + \rho_{\rm r})/3$. If we now change variables to $y \equiv \rho_{\rm m}/\rho_{\rm r} = a/a_{\rm eq}$, we find (after considerable algebra) that

$$\hat{\delta}'' + \frac{2+3y}{2y(1+y)}\hat{\delta}' - \frac{3}{2y(1+y)}\hat{\delta} = 0, \tag{37}$$

  where $\hat{\delta}' \equiv {\rm d}\hat{\delta}/{\rm d}a$.

- If we adopt the trial solution $\hat{\delta} = Cy + D$, then it is easy to demonstrate that this is a solution to the above equation, provided that $D = 2C/3$. Therefore, we can write the growing mode solution as

$$\hat{\delta} = C\left(y + \frac{2}{3}\right), \tag{38}$$

  which becomes independent of y in the limit $y \ll 1$.

- We therefore see that as long as we are in the radiation-dominated regime, our small-scale dark matter perturbations do not grow. Physically, we can understand this effect as follows: the growth rate of the small perturbations (driven by $\rho_{\rm m}$) is much slower than the expansion rate of the Universe (driven by $\rho_{\rm r}$), and so δ is frozen at an approximately constant value for as long as $\rho_{\rm r} \gg \rho_{\rm m}$. Note, however, that this only holds on scales smaller than $\lambda_{\rm J}$. On scales larger than the Jeans length for the radiation-dominated fluid, both $\rho_{\rm r}$ and $\rho_{\rm m}$ contribute to the growth rate of the perturbations, which therefore can still grow significantly during this epoch.

1.5 Peculiar velocities

- Suppose a galaxy is moving with some comoving peculiar velocity $\vec{u}$ relative to the Hubble flow in an otherwise unperturbed universe. In that case, we know from the perturbation equations that

$$\dot{\vec{u}} = -2H\vec{u}. \quad (39)$$

We therefore see that in the absence of pressure forces or gravitational accelerations due to perturbations in the density field, peculiar velocities decay over time. Indeed, since

$$\frac{\mathrm{d}}{\mathrm{d}t} \equiv aH\frac{\mathrm{d}}{\mathrm{d}a}, \quad (40)$$

we can easily show that

$$\frac{\mathrm{d}\ln\vec{u}}{\mathrm{d}\ln a} = -2, \quad (41)$$

and hence that peculiar velocities decay as $u \propto a^{-2}$ and proper velocities decay as $v \propto a^{-1}$ if no forces are acting.

- If we now account for the gravitational acceleration due to the perturbations in the potential associated with the density perturbations, then our equation of motion becomes

$$\dot{\vec{u}} + 2H\vec{u} = \frac{\vec{g}}{a}, \quad (42)$$

where $\vec{g} = -\nabla\delta\Phi/a$. (Note that we are assuming here that we are interested in scales $\lambda \gg \lambda_{\rm J}$, so that pressure forces can still be neglected).

- The general solution to this equation can be decomposed into two parts: one parallel to $\vec{g}$, and a second independent of $\vec{g}$, and which therefore must satisfy

$$\dot{\vec{u}} + 2H\vec{u} = 0. \quad (43)$$

- A guide to the physical interpretation of these solutions is provided by the perturbation equations, which tell us that in the linear regime

$$\nabla \cdot \vec{u} = -\dot{\delta}. \quad (44)$$

This equation shows us that it is possible to have what are known as **vorticity modes**, for which $\nabla \cdot \vec{u} = 0$ and $\dot{\delta}$ vanishes. However, we know that $\dot{\delta}$ is non-zero for any non-zero δ, so these modes are not associated with any perturbation to the density field or the potential. They therefore have no gravitational acceleration associated with them, and hence are the desired solutions of Equation 43.

- We can therefore decompose the peculiar velocity field into a component whose behaviour is governed by the gravitational accelerations induced by the density perturbations, and a second vorticity component that satisfies Equation 43. However, we know that solutions to Equation 43 decay rapidly with time, and so even if we have a vorticity component initially, it will quickly become negligible in comparison to the component induced by the gravitational acceleration due to the density perturbations. It is therefore generally a good approximation to treat the linear velocity field as being curl-free.

- To find the component of the peculiar velocity field parallel to $\vec{g}$, we write it as $\vec{u} = F(t)\vec{g}$. Taking the divergence of this yields

$$\nabla \cdot \vec{u} = F(t) \nabla \cdot \vec{g}. \tag{45}$$

However, we know from the continuity equation that $\nabla \cdot \vec{u} = -\dot{\delta}$, while Poisson's equation in comoving coordinates tells us that $\nabla \cdot \vec{g} = -4\pi G \rho_0 a \delta$. Therefore,

$$\dot{\delta} = 4\pi G \rho_0 a \delta F(t), \tag{46}$$

and so

$$F(t) = \frac{\dot{\delta}}{4\pi G \rho_0 a \delta}. \tag{47}$$

- In a matter-dominated Universe, Friedmann's equation tells us that

$$4\pi G \rho_0 a = \frac{3}{2} H^2 \Omega_{\rm m} a, \tag{48}$$

and so $F(t)$ can also be written as

$$F(t) = \frac{2\dot{\delta}}{3H^2 \Omega_{\rm m} a \delta}. \tag{49}$$

Finally, we can use the fact that

$$\dot{\delta} \equiv \frac{{\rm d}\delta}{{\rm d}t} = aH \frac{{\rm d}\delta}{{\rm d}a} \tag{50}$$

to write $F(t)$ as

$$F(t) = \frac{2 f(\Omega_{\rm m})}{3 H a \Omega_{\rm m}}, \tag{51}$$

where $f(\Omega_{\rm m}) \equiv (a/\delta)({\rm d}\delta/{\rm d}a)$ is a function only of $\Omega_{\rm m}$ and is well-approximated by $f \simeq \Omega_{\rm m}^{0.6}$.

- The comoving peculiar velocity field is therefore given by

$$\vec{u} = \frac{2 f(\Omega_{\rm m})}{3 H a \Omega_{\rm m}} \vec{g}, \tag{52}$$

which in physical units becomes

$$\vec{v} = \frac{2 f(\Omega_{\rm m})}{3 H \Omega_{\rm m}} \vec{g}. \tag{53}$$

- In Fourier space, we can use the fact that $\vec{g}$ and $\vec{k}$ are parallel to derive an expression for the Fourier components of $\vec{v}$ directly from the continuity equation. We find that

$$v_k = -\frac{i H f(\Omega_{\rm m}) a}{k} \hat{\delta}_k, \tag{54}$$

demonstrating that the peculiar velocity field is sensitive to smaller k modes (i.e. larger wavelength modes) than those that dominate the density field.

1.6 Power spectrum

- It is widely believed that the seeds of our density fluctuations were generated by quantum fluctuations occurring during the inflationary epoch. If so, then so long as it remains in the linear regime, the density contrast field δ has a very useful property: it is a homogeneous, isotropic Gaussian random field. Its statistical properties are therefore complete determined by only two numbers: its mean and its variance.

- Mass conservation implies that $\langle \delta \rangle = 0$, where the angle brackets denote a space average.

- The variance of δ is conveniently described in terms of the **power spectrum** $P(k)$:

$$\left\langle \hat{\delta}(\vec{k})\hat{\delta}^{*}(\vec{k}') \right\rangle \equiv (2\pi)^3 P(k) \delta_{\rm D} \left(\vec{k} - \vec{k}' \right), \tag{55}$$

  where $\delta_{\rm D}$ is the Dirac delta function.

- The initial perturbations, seeded by quantum fluctuations during the inflation epoch, are typically assumed to have a power spectrum

$$P_i(k) \propto k, \tag{56}$$

  known as the Harrison-Zel'dovich spectrum.

- However, this initial power spectrum is subsequently modified because perturbations on different scales k do not all grow by the same amount during the radiation-dominated epoch.

- As we saw in the previous section, modes with wavelengths $\lambda \gg r_{\rm H}$ grow as $\delta \propto a^2$ in the radiation-dominated era, and $\delta \propto a$ in the matter-dominated era. On the other hand, modes with $\lambda \ll r_{\rm H}$ do not grow during the radiation-dominated era, and then subsequently begin to grow as $\delta \propto a$ during the matter-dominated era.

- To allow us to examine the effects of this difference in growth rates, let us make two simplifications. We will assume that the behaviour of a given mode changes instantly once $\lambda = r_{\rm H}$, and we will also assume that the evolution of the Universe changes instantly from radiation-dominated to matter-dominated at the redshift of matter-radiation equality (i.e. the redshift at which $\rho_{\rm m} = \rho_{\rm r}$).

- In this simplified picture, modes which have $\lambda > r_{\rm H}$ throughout the radiation-dominated era evolve as $\delta \propto a^2$ throughout the radiation-dominated era, and then as $\delta \propto a$ in the matter-dominated era. On the other hand, modes for which $\lambda = r_{\rm H}$ at some point during the radiation-dominated era evolve initially as $\delta \propto a^2$, then "freeze" once $\lambda = r_{\rm H}$, and finally start to grow again as $\delta \propto a$ at redshifts $z < z_{\rm eq}$. Small-scale modes (with large k) therefore have their growth suppressed relative to large-scale modes (small k).

- To quantify this, we first consider the mode that has $\lambda = r_{\rm H}$ at $z = z_{\rm eq}$; we speak of this mode "entering the horizon" at this time. We can write the comoving wavenumber for this mode as

$$k_0 = a_{\rm eq} \frac{2\pi}{r_{\rm H}} = 2\pi \frac{H_0}{c} \sqrt{\frac{2\Omega_{\rm m,0}}{a_{\rm eq}}} = 2\pi \frac{H_0}{c} \Omega_{\rm m,0} \sqrt{\frac{2}{\Omega_{\rm r,0}}}. \quad (57)$$

- Next, consider some mode that enters the horizon at the point when the scale factor is $a_{\rm enter} < a_{\rm eq}$. Up to this point, this mode has grown at the same rate as the mode with wavenumber k_0, but during the period from $a_{\rm enter}$ to $a_{\rm eq}$, it does not grow. On the other hand, the larger mode continues to grow as $\delta \propto a^2$ during this period.

- At $a_{\rm eq}$, the smaller mode is therefore suppressed relative to the larger mode by a factor

$$f_{\rm sup} = \left(\frac{a_{\rm enter}}{a_{\rm eq}}\right)^2 = \left(\frac{k_0}{k}\right)^2. \quad (58)$$

- After we enter the matter-dominated regime, the relative size of the modes does not change (so long as we remain in the linear regime). Since the power spectrum scales as δ^2, the final power spectrum is therefore related to the initial power spectrum by:

$$P_f(k) \propto \begin{cases} f_{\rm sup}^2 P_i(k) & k > k_0 \\ P_i(k) & k < k_0 \end{cases} \quad (59)$$

- If our initial power spectrum is the Harrison-Zel'dovich spectrum, we find that

$$P_f(k) \propto \begin{cases} k^{-3} & k > k_0 \\ k & k < k_0 \end{cases} \quad (60)$$

  where we have made use of the fact that $f_{\rm sup} \propto k^{-2}$.

- This behaviour of the power spectrum has important consequences when we come to consider the formation of highly non-linear structures.

1.7 Relative velocity of dark matter and baryons

- Prior to recombination, the baryons and the radiation are tightly coupled together by Compton scattering, which allows for efficient momentum transfer from one component to another.

- As already noted, an important consequence of this is that the effective sound-speed in this coupled fluid is very high: $c_{\rm s,eff} = c/\sqrt{3}$, where c is the speed of light.

- Another important consequence is the fact that small-scale perturbations in the baryonic component are smoothed away by an effect known as **Silk damping**.

- If we have an overdensity, then locally we will have a higher number density of photons than in the surrounding gas. These photons will try to diffuse away from the overdensity, in order to restore the photon number density to equilibrium. Because of the high optical depth of the Universe at this epoch, they will do this via radiation diffusion (i.e. each photon will execute a random walk away from its initial location). As they do so, they will drag the baryons along with them, owing to the strong momentum coupling between baryons and photons.

- We can write the photon mean free path as

$$\lambda_{\rm mfp} = \frac{1}{n_{\rm e}\sigma_{\rm T}}, \tag{61}$$

where $\sigma_{\rm T}$ is the Thomson scattering cross-section. The diffusion coefficient is then given by

$$D = \frac{1}{3}\lambda_{\rm mfp}c, \tag{62}$$

and the diffusion radius (i.e. the distance to which the photons diffuse in time t) is given by

$$r_{\rm D} \simeq \sqrt{Dt}. \tag{63}$$

- At recombination, $t \sim 10^{13}$ s and $n_{\rm e} \simeq 400\ {\rm cm}^{-3}$. Therefore, $\lambda_{\rm mfp} \simeq 1.2$ kpc and $r_{\rm D} \simeq 6.2$ kpc, where these distances are in *physical units*. In comoving units, the diffusion length corresponds to ~ 6 Mpc, and hence Silk damping will erase any perturbations in the baryon-photon fluid on scales smaller than this.

- On scales $r > r_{\rm D}$, perturbations survive. As we have seen, we can consider the linear perturbations on these larger scales to be built up of a superposition of sound waves. Detailed analysis of the behaviour of the perturbations in this regime shows that owing to the effects of constructive interference, we expect to get the largest effects on wavelengths that are harmonics of the horizon scale, i.e. $\lambda = \frac{1}{n}\frac{c}{H(z)}$, where n is an integer, provided that $\lambda > r_{\rm D}$.

- This is a strong prediction of the basic hot Big Bang model, and has been successfully confirmed – these so-called "acoustic oscillations" are responsible for oscillatory pattern that we see if we measure the strength of the CMB anisotropies on a range of different angular scales, and we will discuss them in more detail later in this course.

- Now, what happens once the Universe recombines? Clearly, $n_{\rm e}$ drops rapidly and hence the photon mean free path increases significantly. However, at the same time, the coupling between photons and baryons becomes much weaker, as the timescale on which the two components can exchange momentum becomes comparable to or greater than the expansion timescale. Therefore, at a redshift $z \sim 1000$, the photons and baryons **decouple**. Although some scattering events occur after this time, and there remains a transfer of energy from the photons to the baryons, the rate at which momentum is transferred becomes too small to significantly affect the mean momentum of the baryons, and perturbations in the photons and in the baryons no longer evolve in the same fashion.

- As a result, the sound speed of the baryons drops very sharply from $c/\sqrt{3}$ to $c_{\rm s} = \sqrt{kT/\mu m_{\rm H}}$, the usual thermal sound-speed of an ideal gas. The Jeans length in the baryons also drops sharply, and on small-scales the baryons start to fall into the small-scale potential wells created by the dark matter. The dark matter, of course, does not couple to the radiation, and hence the perturbations in this component are not affected by Silk damping. Therefore, the small-scale perturbations in the baryons are regenerated, thanks to the dark matter, while the radiation component remains smooth on these scales.

- All of the effects that I have described so far were understood by the late 70s and early 80s. However, in 2010, Tseliakovich & Hirata pointed out another consequence of the baryon-photon coupling that had previously been overlooked. Before decoupling, the baryon-photon fluid has a non-zero velocity relative to the dark matter, owing to the effect of the acoustic oscillations in the former. What Tseliakovich & Hirata realized was that the baryons would initially retain this relative velocity even after decoupling.

- Detailed calculations (e.g. Tseliakovich & Hirata, 2010, Phys. Rev. D, 82, 083520) show that at decoupling, the rms size of the relative velocity[2] is around $30\,{\rm km\,s^{-1}}$. This is very small compared to the sound-speed prior to decoupling, but is large compared to the sound-speed of the baryons after decoupling, which is ~ 5–$6\,{\rm km\,s^{-1}}$.

- The coherence length of this relative velocity is comparable to the Silk damping scale, i.e. a few comoving Mpc. On small scales, therefore, the motion of the gas relative to the dark matter can be modelled as a bulk velocity. The size of this velocity decreases as the Universe expands – as with any peculiar velocity, it falls off as $v_{\rm pec} \sim (1+z)$. However, the sound speed in the gas also falls off with decreasing redshift, initially as $c_{\rm s} \propto (1+z)^{1/2}$ in the regime where $T_{\rm gas} \simeq T_{\rm r}$, and then as $c_{\rm s} \propto (1+z)$ in the regime where $T_{\rm gas}$ evolves adiabatically.

- At $z \sim 100$ – approximately the redshift at which the behaviour of T changes – the rms streaming velocity is around $3\,{\rm km\,s^{-1}}$ and the sound-speed is around $1.7\,{\rm km\,s^{-1}}$, and so the streaming motions are still supersonic. They remain so at lower redshift, as from this point on both $c_{\rm s}$ and $v_{\rm pec}$ evolve similarly with redshift.

- The full effects of this bulk motion on the formation of structure remain to be explored, but one obvious effect will be to increase the effective Jeans mass of the gas by a factor

$$f_{\rm inc} = \left(\frac{v_{\rm pec}}{c_{\rm s}}\right)^3 \sim 10. \tag{64}$$

[2]Note that in a homogeneous, isotropic Universe, the *mean* streaming velocity must be zero, but the *root-mean-squared* (rms) streaming velocity need not be zero.

2 Formation of structure: non-linear regime

2.1 The spherical collapse model

- Our treatment above works well in the linear regime, when $|\delta| \ll 1$, but breaks down once $|\delta| \sim 1$, since at this point we are no longer dealing with small perturbations, and hence can no longer use the tools of linear perturbation theory.

- The evolution of the gas and dark matter in the so-called **non-linear** regime is very complicated, and in general we need to use numerical simulations, rather than analytical techniques, in order to follow it.

- However, there are a few useful approximate models that we can look at that give us some guidance as to the behaviour of the gas and dark matter in the non-linear regime.

- The first example that we're going to look at is known as the **spherical collapse** model.

- Consider a spherical overdensity with uniform internal density. As this perturbation is overdense, it will reach some maximum physical radius and then collapse due to its own self-gravity. We denote the metric scale-factor at which the perturbation reaches its turn-around radius as $a_{\rm ta}$, and the radius of the perturbation at this point as $R_{\rm ta}$. We then define dimensionless coordinates:
$$x \equiv \frac{a}{a_{\rm ta}}, \qquad y \equiv \frac{R}{R_{\rm ta}}. \tag{65}$$

- If we consider, for simplicity, an Einstein-de Sitter Universe, then we can write the Friedmann equation as
$$\frac{{\rm d}x}{{\rm d}\tau} = x^{-1/2}, \tag{66}$$
where $\tau \equiv H_{\rm ta} t$ and $H_{\rm ta} = H_0 a_{\rm ta}^{-3/2}$.

- The equation of motion for the radius of our sphere can be written as
$$\ddot{R} = -\frac{GM}{R^2}, \tag{67}$$
$$= -\frac{4\pi}{3}\rho_{\rm ta} R_{\rm ta}^3 \frac{G}{R^2}. \tag{68}$$
Converting from t to τ, and defining a new overdensity parameter ζ through the equation
$$\rho_{\rm ta} = \frac{3H_{\rm ta}^2}{8\pi G}\zeta \tag{69}$$
allows us to write this in a much simpler form:
$$\frac{{\rm d}^2 y}{{\rm d}\tau^2} = -\frac{\zeta}{2y^2}. \tag{70}$$
Note that ζ is simply the overdensity of our perturbation at turn-around with respect to the cosmological background at the same time, measured in units of $\rho_{\rm crit}$.

- To solve our equation of motion, we need to specify some boundary conditions. The obvious choices are
$$\left.\frac{\mathrm{d}y}{\mathrm{d}\tau}\right|_{x=1} = 0, \qquad y|_{x=0} = 0, \tag{71}$$
i.e. our perturbation starts with zero radius when $a = 0$ and reaches its maximum size when $a = a_{\mathrm{ta}}$.

- With these boundary conditions, and with the help of the Friedmann equation, we can obtain the following solution
$$\tau = \frac{1}{\sqrt{\zeta}}\left[\frac{1}{2}\arcsin(2y-1) - \sqrt{y-y^2} + \frac{\pi}{4}\right], \tag{72}$$
which cannot easily be inverted to give y in terms of τ.

- At turn-around, $x = y = 1$ and $\tau = 2/3$, which means that
$$\zeta = \left(\frac{3\pi}{4}\right)^2 \simeq 5.55. \tag{73}$$

- By symmetry, the time taken from turn-around to collapse must be the same as that taken from the start to turn-around, i.e. in the absence of pressure forces or any non-sphericity, our perturbation will collapse to a point at $\tau = 4/3$, corresponding to $x = 2^{2/3}$.

- If our perturbation had not begun to evolve non-linearly, but had simply continued to evolve following the linear solution, its overdensity at this point would be merely
$$\delta_{\mathrm{c}} = 2^{2/3}\delta_{\mathrm{ta}} \simeq 1.69. \tag{74}$$

- In reality, our perturbation will never be perfectly spherical; non-spherical motions will develop as the perturbation collapses and will eventually halt the collapse.

- We assume that after the collapse halts, the collapsed object – often referred to as a **dark matter halo**, assuming we're considering a perturbation in the dark matter – relaxes into a state of **virial equilibrium**. In this case, the virial theorem tells us that the magnitude of the potential energy of the halo is equal to twice its kinetic energy:
$$|W_{\mathrm{vir}}| = 2T_{\mathrm{vir}} \tag{75}$$
Energy conservation implies that the kinetic energy of the virialized halo must be equal to the difference between the potential energy at turnaround, W_{ta}, and the potential energy of the virialized halo:
$$|W_{\mathrm{vir}}| - |W_{\mathrm{ta}}| = T_{\mathrm{vir}}. \tag{76}$$
Therefore,
$$|W_{\mathrm{ta}}| = T_{\mathrm{vir}}, \tag{77}$$
$$|W_{\mathrm{vir}}| = 2|W_{\mathrm{ta}}|. \tag{78}$$

Since the potential energy of a spherical perturbation of radius R scales as $1/R$, this implies that

$$R_{\rm vir} = \frac{R_{\rm ta}}{2}. \tag{79}$$

- We can use this result to solve for the overdensity of the perturbation with respect to the background density at the time that the collapsing perturbation first virializes. Two factors contribute to this: the perturbation has collapsed (and hence increased its density), and the Universe has expanded (and hence decreased its density). The resulting density contrast is given by

$$\Delta = \left(\frac{2^{2/3}}{1/2}\right)^3 \zeta = 32\zeta = 18\pi^2 \simeq 178. \tag{80}$$

- Up to this point, we have been assuming an Einstein-de Sitter cosmological model. A similar analysis in the case where $\Omega_{\rm m} \neq 1$ is possible, but requires us to solve the resulting equations numerically. However, the end result is not too different from the Einstein-de Sitter case. For example, for $\Omega_{\rm m,0} = 0.3$ and $\Omega_\Lambda = 0.7$, we find that at $z = 0$, $\Delta \simeq 100$.

- In reality, non-linear structures forming in the dark matter are unlikely to be perfectly spherical. Indeed, **N-body** simulations that model the full non-linear evolution of the dark matter (albeit with some finite mass resolution) show that much is located in mildly overdense filaments and sheets, with larger overdensities located within these structures, particularly at the intersection of filaments.

- These highly overdense regions typically have an ellipsoidal morphology, and are commonly referred to as **dark matter halos**. Halos that have masses that exceed the local effective Jeans mass of the gas can capture gas from their surroundings. If this gas then cools and undergoes further gravitational collapse, then the formation of stars will be the end result. In other words, these dark matter halos are the locations in which galaxies form. It is therefore important to understand their properties and their abundance within the Universe.

- In practice, even though these dark matter halos are far less symmetric than the idealized perturbation that we have considered in this section, the results of the spherical collapse model provide a reasonable first approximation when discussing their properties. This simple model also gives us a basis for determining the number density of halos of a given mass that we expect to find in the Universe, as we will see later.

2.2 Halo density profiles

- Although the structure of simulated dark matter halos can be complex when looked at in detail, there are some surprising underlying regularities. Most prominent amongst these is the **halo density profile**.

- If we spherically average the dark matter density (i.e. compute the mean density of dark matter in a series of spherical shells that lie at increasing radial distances r from the centre of mass), then we find that the density profile is well approximated at most radii by the **Navarro-Frenk-White** or NFW profile:

$$\rho(r) = \frac{\delta_{\rm c}}{(r/r_{\rm s})(1+r/r_{\rm s})^2}\rho_0, \tag{81}$$

where ρ_0 is the cosmological background density.

- These profile is characterized by two numbers: a scale radius $r_{\rm s}$ and a characteristic overdensity $\delta_{\rm c}$. At distances $r \ll r_{\rm s}$, the NFW profile scales as $\rho(r) \propto r^{-1}$, while at $r \gg r_{\rm s}$ the profile steepens, falling off as $\rho(r) \propto r^{-3}$.

- The fact that $\rho \rightarrow \infty$ as $r \rightarrow 0$ tells us that this description of the density profile must break down at some point, and indeed we expect this to occur on scales comparable to the **free-streaming length** of the dark matter particles, i.e. the mean distance that the particles move due to their random velocities within one dynamical time. However, for cold dark matter, this length-scale is small, and hence it is believed that the NFW profile (or something qualitatively similar) remains a good description of the dark matter until r is very small. CDM halos are therefore believed to have strong **density cusps**.

- We can compute the mass associated with a given dark matter halo simply by integrating the density profile out to the edge of the halo:

$$M(r) = 4\pi \int_0^{r_{\rm h}} r^2 \rho(r) {\rm d}r = 4\pi\delta_{\rm c}\rho_0 r_{\rm s}^3 \left[\ln(1+x) - \frac{x}{1+x}\right], \tag{82}$$

where $x = r_{\rm h}/r_{\rm s}$. However, this prompts the question of how to define the "edge" of the halo.

- The dark matter density distribution within the halo connects smoothly to its surroundings, without any sudden discontinuity marking a real, physical edge. Therefore, to some extent, the choice of an edge, and hence of a mass for the halo, is a matter of convention.

- Typically, we attempt to associated the edge of the dark matter halo with its virial radius, and we estimate the size of this in a way that is inspired by the spherical collapse model discussed in the previous section. We saw previously that in an Einstein-de Sitter universe, a virialized spherical perturbation has a mean density that is $18\pi^2 \simeq 178$ times higher than the cosmological background density at the moment of virialization.

- The virial radius of a halo is therefore often defined to be the radius in the spherically-averaged density profile enclosing some specified mean overdensity. The spherical collapse model suggests that the overdensity used in this definition should be $18\pi^2$, but for simplicity it is common to instead take a value of 200. We therefore associate the "edge" of the halo with the radius r_{200} for which

$$M(r_{200}) \equiv M_{200} = \frac{4\pi}{3} r_{200}^3 \times 200\rho_0. \tag{83}$$

- Using our expression for $M(r)$, it is easy to show that M_{200} can be written as The associated virial mass, M_{200}, is then simply given by

$$M_{200} = 4\pi\delta_{\rm c}\rho_0 r_{\rm s}^3 \left[\ln(1+c) - \frac{c}{1+c}\right], \tag{84}$$

where $c = r_{200}/r_{\rm s}$ is a parameter known as the **concentration** of the halo. We therefore have

$$4\pi\delta_{\rm c}\rho_0 r_{\rm s}^3 \left[\ln(1+c) - \frac{c}{1+c}\right] = \frac{4\pi}{3} r_{200}^3 \times 200\rho_0. \tag{85}$$

Rearranging this, we find that $\delta_{\rm c}$ is related to the concentration by

$$\delta_{\rm c} = \frac{200}{3} \frac{c^3}{[\ln(1+c) - c/1+c]}. \tag{86}$$

- We therefore see that instead of specifying $r_{\rm s}$ and $\delta_{\rm c}$, we can instead fully characterize the NFW density profile of a halo by specifying c and M_{200}. Furthermore, we find empirically that c is only a weakly varying function of the halo mass and the redshift of formation of the halo. (See e.g. Ludlow et al., 2014, MNRAS, 441, 378).

- It remains unclear exactly why the spherically-averaged density profiles of dark matter halos are approximated so well by the NFW profile. It presumably has something to do with the physics of halo assembly, but this is a topic that is still not well understood.

- It is also unclear whether the dark matter halos associated with real galaxies actually follow the NFW profile. In particular, efforts to constrain the dark matter density profile close to the centre of dark matter dominated dwarf spheroidal galaxies tend to indicate a profile with a flat, constant density core, rather than a power-law cusp. This result is in direct contradiction to the results of simulations of halo growth that follow only the dark matter, but it has recently been suggested that the inclusion of the effects of stellar feedback may resolve this difficulty (e.g. Pontzen & Governato, 2012, MNRAS, 421, 3464).

2.3 The Zeldovich approximation

- The Zeldovich approximation is a simple but powerful model for the evolution of density perturbations in the non-linear regime. At heart, it is a *kinematical* approach to the growth of structure: we work out the initial displacement of particles due to some perturbation and then simply assume that they continue to move in the same direction at all later times.

- To express this mathematically, we write the proper coordinates $\vec{x}$ of a given particle as

$$\vec{x}(t) = a(t)\vec{q} + b(t)\vec{f}(\vec{q}). \tag{87}$$

If $b(t) \to 0$ as $t \to 0$, then the second term vanishes as we approach $t = 0$ and we simply have the usual mapping between proper and comoving coordinates. We therefore identify $\vec{q}$ at $t = 0$ as the initial comoving coordinate of the particle in question.

- In the absence of the second term, it is easy to see that the subsequent evolution of $\vec{x}$ would simply be given by the Hubble flow. The effect of the second term is therefore to perturb the position of our particle away from where it would be if it simply followed the Hubble flow – i.e. it corresponds to the displacement due to the peculiar velocity field.

- The key assumption at the heart of the Zeldovich approximation is that this displacement can be written in the form that we have used above, as the product of a time-independent **displacement field** $\vec{f}(\vec{q})$ and a time-dependent scale factor $b(t)$.

- In order for this approximation to be useful, we need some way to relate b and $\vec{f}$ to other quantities that we already know. We start by relating the density perturbation to the displacement field $\vec{f}$.

- Mass conservation requires that the mass associated with any small Lagrangian volume (i.e. a volume moving with the flow) remains constant. Therefore, the density at position $\vec{x}$ at time t, $\rho(\vec{x}, t)$ is related to the initial density $\rho_{\rm init}(\vec{q}, 0)$ by

$$\rho = \rho_{\rm init} \left| \frac{\partial \vec{x}}{\partial \vec{q}} \right|, \tag{88}$$

where $|\partial \vec{x} / \partial \vec{q}|$ is the Jacobian determinant. This expression can be rewritten in terms of the cosmological background density ρ_0 as

$$\rho = \rho_0 \frac{1}{|\delta_{ij} - (b/a) \partial f_i / \partial q_j|}. \tag{89}$$

- If the displacement field is the result of growing density perturbations, then we know from our previous discussion of the peculiar velocity field that it must be irrotational. Therefore, we can write it as the gradient of some scalar potential:

$$\vec{f}(\vec{q}) = \nabla \psi(\vec{q}). \tag{90}$$

It then follows that

$$\frac{\partial f_i}{\partial q_j} = \frac{\partial^2 \psi}{\partial q_i \partial q_j}. \tag{91}$$

Therefore, the tensor defined by $\partial f_i / \partial q_j$, known as the **strain tensor** or **deformation tensor** is symmetric, allowing us to diagonalize it. After doing so, we find that

$$\frac{\rho}{\rho_0} = \left[\left(1 - \frac{b}{a}\alpha\right) \left(1 - \frac{b}{a}\beta\right) \left(1 - \frac{b}{a}\gamma\right) \right]^{-1}, \tag{92}$$

where $-\alpha$, $-\beta$ and $-\gamma$ are the three eigenvalues of the deformation tensor.

- We therefore see that in the general case where α, β and γ all have different values, collapse occurs first along the principal axis corresponding to the largest eigenvalue. In other words, if α is the largest eigenvalue, $\rho \rightarrow \infty$ as $(b/a)\alpha \rightarrow 1$. The evolving density field therefore first makes sheet-like structures, known as **Zeldovich pancakes**, that then fragment into filaments and spherical halos.

- In the linear regime, where the terms involving α, β and γ are small, Equation 92 simplifies to

$$\frac{\rho}{\rho_0} = 1 + \frac{b}{a}(\alpha + \beta + \gamma). \tag{93}$$

Since $1 + \delta = \rho/\rho_0$, it follows that

$$\delta = \frac{b}{a}(\alpha + \beta + \gamma) = -\frac{b}{a}\nabla \cdot \vec{f}. \tag{94}$$

- We can also show, based on our definition of $\vec{x}$, that the comoving velocity perturbation is given by

$$\vec{u} = \frac{1}{a}\left(\dot{\vec{x}} - H\vec{x}\right) = \left(\frac{\dot{b}}{a} - \frac{\dot{a}b}{a^2}\right)\vec{f}. \tag{95}$$

With a little additional algebra one can then verify that δ and $\vec{u}$ satisfy the mass conservation equation $\nabla \cdot \vec{u} = -\dot{\delta}$, as we expect for linear perturbations.

- From Equation 94, we see that the evolution of δ with time is given by the function $b(t)/a(t)$, since the displacement field $\vec{f}$ is independent of time. However, we already know that the evolution of δ with time is given by the linear growth factor $D_+(t)$. We therefore see immediately that

$$\frac{b(t)}{a(t)} = D_+(t). \tag{96}$$

- Using this expression for $b(t)$, we can then show that

$$\vec{u} = \frac{\dot{\delta}}{\delta}\vec{f}, \tag{97}$$

$$= f(\Omega_{\rm m})H\vec{f}, \tag{98}$$

where $f(\Omega_{\rm m}) = (a/\delta){\rm d}\delta/{\rm d}a$ as before. However, we also know that in the linear regime,

$$\vec{u} = \frac{2f(\Omega_{\rm m})}{3Ha\Omega_{\rm m}}\vec{g}. \tag{99}$$

We therefore see that

$$\vec{f} = \frac{2}{3H^2\Omega_{\rm m}}\frac{\vec{g}}{a}, \tag{100}$$

or in other words that our required displacement field is directly proportional to the initial gravitational acceleration.

- This relationship allows us to understand why the Zeldovich approximation works very well for describing the formation of pancakes. Consider a flow of matter towards a dense sheet with infinite extent perpendicular to the flow. The gravitational acceleration due to this sheet is independent of our distance from it, and hence remains constant. Therefore, the motion of the matter at early times, when the density perturbation is small, remains a good predictor of its behaviour at late times, even after we are no longer in the linear regime.

- Once the inflowing streams of matter intersect with each other, the Zeldovich approximation breaks down. Formally, it predicts that the density should become infinite at this point, whereas we know in reality that short-range gravitational interactions that are not treated in the approximation will become dominant. Nevertheless, prior to this event (often referred to as **shell-crossing**) it remains a useful guide to the behaviour of the matter.

2.4 The Press-Schechter mass function

- Ideally, we would like to be able to determine the number density of halos of a given mass – the halo **mass function** – as a function of redshift without going to all the trouble and expense of running a large N-body simulation.

- Fortunately, we can! There is a simple analytical argument that allows us to derive a mass function that turns out to be a reasonable approximation to the true mass function. This argument was first formulated by Press & Schechter in 1974, and the resulting mass function has become known as the **Press-Schecher mass function**.

- We start by assigning a length scale $R(M)$ to each halo of mass M via

$$R(M) = \left(\frac{3M}{4\pi\rho_{\rm cr}(z)\Omega_{\rm m}(z)} \right)^{1/3}. \tag{101}$$

  (In other words, R is the radius of a uniform sphere filled with matter at the mean density that has a total mass M).

- We next consider the density contrast smoothed on this scale R. This is defined as

$$\bar{\delta}_R(\vec{x}) \equiv \int {\rm d}^3 y \delta(\vec{x}) W_R(\vec{x} - \vec{y}), \tag{102}$$

  where $W_R(\vec{x} - \vec{y})$ is a suitably chosen **window function**.

- If the density contrast δ is a Gaussian random field, then so is the smoothed field $\bar{\delta}_R$. For a Gaussian random field, the probability of finding any particular value $\bar{\delta}$ at a point in space $\vec{x}$ is given by

$$p(\bar{\delta}) = \frac{1}{\sqrt{2\pi\sigma_R^2}} \exp\left[-\frac{\bar{\delta}^2(\vec{x})}{2\sigma_R^2} \right], \tag{103}$$

where σ_R^2 is the smoothed density variance

$$\sigma_R^2 = 4\pi \int_0^\infty \frac{k^2 \mathrm{d}k}{(2\pi)^3} P(k) \hat{W}_R^2(k), \tag{104}$$

and $\hat{W}_R$ is the Fourier transform of our window function.

- The fraction of all points that have a density contrast greater than δ_c (the linear density contrast for spherical collapse) is then given by

$$F = \int_{\delta_\mathrm{c}}^\infty p(\bar{\delta})\,\mathrm{d}\bar{\delta}, \tag{105}$$

$$= \frac{1}{2}\mathrm{erfc}\left(\frac{\delta_\mathrm{c}}{\sqrt{2}\sigma_R}\right), \tag{106}$$

where erfc is the complementary error function.

- The great insight of Press & Schechter was that this number could also be identified as the total mass fraction in halos of masses greater than or equal to M.

- Another way of thinking about this: in the unsmoothed linear density contrast field, any points that have $\delta > \delta_\mathrm{c}$ correspond to gas that is now in a collapsed structure. By smoothing the density contrast field, we filter out those points that are in structures with scales less than $R(M)$ or masses less than M; hence, whatever is left must be in structures with mass $\geq M$.

- The mass fraction in halos with masses in the range $M, M + \mathrm{d}M$ is simply $\partial F/\partial M$. To compute this, we use the fact that we can write $\partial/\partial M$ as

$$\frac{\partial}{\partial M} = \frac{\mathrm{d}\sigma_R}{\mathrm{d}M}\frac{\partial}{\partial \sigma_R}, \tag{107}$$

and also use the identity

$$\frac{\mathrm{d}}{\mathrm{d}x}\mathrm{erfc}(x) \equiv -\frac{2}{\sqrt{\pi}} e^{-x^2}. \tag{108}$$

We find

$$\frac{\partial F}{\partial M} = \frac{1}{\sqrt{2\pi}} \frac{\delta_\mathrm{c}}{\sigma_R} \frac{\mathrm{d}\ln\sigma_R}{\mathrm{d}M} \exp\left(-\frac{\delta_\mathrm{c}^2}{2\sigma_R^2}\right). \tag{109}$$

- If we integrate this over all masses, we find we have a normalization problem:

$$\int_0^\infty \frac{\partial F(M)}{\partial M}\mathrm{d}M = \frac{1}{2}. \tag{110}$$

Press & Schechter dealt with this by (somewhat arbitrarily) multiplying the mass function by a factor of two. The actual resolution to this problem was recognized 17 years later by Bond et al. (1991, ApJ, 379, 440), and requires us to derive the mass function in a somewhat different fashion, using the methods of excursion set theory. However, this is outside the scope of the present course.

- Given the correctly normalized version of $\partial F/\partial M$, we can then compute the comoving halo number density simply by multiplying by ρ_0/M:

$$N(M,z)\,\mathrm{d}M = \sqrt{\frac{2}{\pi}}\frac{\rho_0\delta_\mathrm{c}}{\sigma_R}\frac{\mathrm{d}\ln\sigma_R}{\mathrm{d}M}\exp\left(-\frac{\delta_\mathrm{c}^2}{2\sigma_R^2}\right)\frac{\mathrm{d}M}{M}. \tag{111}$$

- The redshift dependence of this expression enters because σ_R increases as the Universe expands and the density perturbations grow. It is therefore often convenient to write the above Equation in terms of $\sigma_{\mathrm{R},0}$, the variance of the linear density field at $z=0$, and the linear growth factor $D_+(z)$. In this case, we have

$$N(M,z)\,\mathrm{d}M = \sqrt{\frac{2}{\pi}}\frac{\rho_0\delta_\mathrm{c}}{D_+(z)\sigma_{R,0}}\frac{\mathrm{d}\ln\sigma_{R,0}}{\mathrm{d}M}\exp\left(-\frac{\delta_\mathrm{c}^2}{2D_+(z)^2\sigma_{R,0}^2}\right)\frac{\mathrm{d}M}{M}. \tag{112}$$

- To help us understand the behaviour of this mass function, let us start by considering the simple case in which our power spectrum $P(k)$ is a power-law function of k, i.e. $P(k)\propto k^n$. In this case, $\sigma_{\mathrm{R},0}$ is given by

$$\sigma_{R,0}^2 = 4\pi\sigma_N^2\int_0^\infty \frac{k^{2+n}\mathrm{d}k}{(2\pi)^3}\hat{W}_R^2(k), \tag{113}$$

where σ_N is some appropriately chosen normalization factor that fixes the normalization of the power spectrum. We often choose to express this normalization in terms of σ_8, the value of σ at $z=0$ within a sphere of radius $R=8h^{-1}$Mpc.

- If we assume, for simplicity, that our window function is a top-hat in k-space, so that

$$\hat{W}_R = \begin{cases} 0 & k > 2\pi/R \\ 1 & k < 2\pi/R \end{cases} \tag{114}$$

then we find that

$$\sigma_{R,0}^2 \propto \int_0^{2\pi/R} k^{2+n}\mathrm{d}k \propto R^{-3+n}. \tag{115}$$

Since $R\propto M^{1/3}$, we therefore find that $\sigma_{R,0}\propto M^{-(3+n)/6}$.

- If we consider small scales, so that we can set the exponential term in our mass function equal to one, then we find that

$$N(M,z)\,\mathrm{d}M \propto M^{(n-9)/6}\mathrm{d}M. \tag{116}$$

We saw in a previous lecture that $P(k)\propto k^{-3}$, and hence on small scales $n=-3$. We therefore find that at the low-mass end, the mass function scales as

$$N(M,z)\,\mathrm{d}M \propto M^{-2}\mathrm{d}M. \tag{117}$$

- We therefore see that there are many more low-mass halos than high-mass halos. Moreover, the mass found in each logarithmic mass bin is constant, demonstrating that these low-mass halos do not only dominate the number counts but also represent a significant fraction of the total available mass.

- At the high mass end of the mass function, the exponential term generally dominates. The presence of this term means that although, in principle, there is a non-zero probability of finding a halo of arbitrarily large mass at any given redshift, in practice the probability soon becomes so small that the chance of finding one within the observable Universe is tiny; i.e. we may as well consider it to be zero, for all intents and purposes.

- It is often useful to quantify the rarity of a given halo in terms of the argument of this exponential. For instance, suppose that we are interested in a halo with a mass such that

$$\frac{\delta_{\rm c}}{\sigma_{R,0} D_+(z)} = 3. \tag{118}$$

Rearranging this expression, we find that

$$\sigma_{R,0} = \frac{1}{3} \frac{\delta_{\rm c}}{D_+(z)}, \tag{119}$$

and hence in order to form such a halo, we need a local upwards fluctuation in the density contrast field that corresponds to a three-sigma fluctuation. We know from numerical integration of the Gaussian distribution that such a fluctuation occurs with a probability of around 1%, and hence around 1% of the total mass in the Universe is to be found in regions where δ is this large or larger.

3 Gravitational lensing

- Because of space-time curvature, concentrations of mass (or other forms of energy) deflect light towards themselves, a phenomenon known as **gravitational lensing**.

- Lensing comes in two main varieties: **strong lensing**, when the deflection is strong enough to give rise to multiple background sources, and **weak lensing**, when there is only a single image produced, albeit with some distortion of shape and change of brightness due to the lensing.

- The key quantity that determines whether a gravitational lens (i.e. a foreground matter concentration such as a galaxy or cluster) causes strong or weak lensing is its projected mass surface density Σ. If Σ exceeds a critical surface density

$$\Sigma_{\rm crit} = \frac{c^2}{4\pi G} \frac{D_{\rm s}}{D_{\rm d} D_{\rm ds}}, \tag{120}$$

where $D_{\rm s}$ is the angular diameter distance from the observer to the source, $D_{\rm d}$ is the angular diameter distance from the observer to the lens and $D_{\rm ds}$ is the angular diameter distance from the lens to the source, then we have strong lensing; otherwise, we have weak lensing.

- For lenses sources at cosmological distances, the effective distance

$$D_{\rm eff} \equiv \frac{D_{\rm d} D_{\rm ds}}{D_{\rm s}} \tag{121}$$

is of the order of a Gpc or more, and the critical density is

$$\begin{aligned} \Sigma_{\rm cr} &\simeq 0.35 \left(\frac{D_{\rm eff}}{1\ {\rm Gpc}}\right)^{-1} {\rm g\ cm^{-2}}, \\ &\simeq 1700 \left(\frac{D_{\rm eff}}{1\ {\rm Gpc}}\right)^{-1} {\rm M_\odot\ pc^{-2}}. \end{aligned} \tag{122}$$

- Surface densities greater than $\Sigma_{\rm cr}$ are typically found only in highly overdense structures, such as galaxy clusters. Therefore, although strong lensing can be useful for learning about the properties of these structures, it is of little use for learning about the large-scale structure of the Universe. This is the domain of weak lensing, and it is on this variety of lensing that we will focus for the rest of this lecture.

3.1 Deflection angle, convergence and shear

- In the weak lensing regime, the Newtonian gravitational potential is small ($\Phi \ll c^2$) and it is possible to show that the effect of lensing by a mass distribution between us and some distant source can be described by the following equation

$$\vec{\alpha}(\vec{\theta}) = \frac{2}{c^2}\int_0^r {\rm d}r' \frac{f_K(r-r')}{f_K(r)} \nabla_\perp \Phi[f_{\rm K}(r')\vec{\theta}]. \tag{123}$$

Here $\vec{\theta}$ is the angular position on the sky and $\vec{\alpha}$ is the **deflection angle** corresponding to that position on the sky, i.e. the angular amount by which the light-ray observed at those angular coordinates has been deflected during its travel between the distant source and the observer. In the integrand, $f_K(r)$ is the comoving angular diameter distance, given by

$$f_K(r) = \begin{cases} \sin(r) & K=1 \\ r & K=0 \\ \sinh(r) & K=-1 \end{cases} \tag{124}$$

where r is our comoving radial coordinate. Finally, $\nabla_\perp\Phi$ is the gradient of the Newtonian gravitational potential in a direction perpendicular to the path of the light-ray.

- The main features of this equation are easy to understand. The deflection is sensitive only to the component of $\nabla\Phi$ perpendicular to the ray because the component parallel to the ray may change the energy of the associated photons (via gravitational redshifts or blueshifts) but will not change their direction of propagation. The sensitivity of $\vec{\alpha}$ to the potential gradient is weighted by ratio of the angular diameter distance from mass to source with that from observer to mass because mass concentrations close to the observer give rise to large deflections, while those close to the source give rise to small deflections.

- Now, it is easy to see that the deflection angle α itself is not an observable: if all of the light-rays coming from a distant object were deflected by the same amount, all that would happen would be that its apparent angular position on the sky would be different from its position in the absence of lensing. However, in this case we have no way of knowing what its angular position would have been in the absence of lensing, and hence no way of measuring $\vec{\alpha}$.

- From this, it should be clear that what we are actually sensitive to are *changes* in $\vec{\alpha}$ from one light-ray to another, as it is these that lead to changes in the images properties compared to the no-deflection case. A key quantity in this context is the derivative of the deflection angle with respect to the position on the sky, which is a 2×2 matrix with components

$$\frac{\partial \alpha_i}{\partial \theta_j} = \frac{2}{c^2} \int_0^r \mathrm{d}r' \frac{f_K(r-r') f_K(r')}{f_K(r)} \frac{\partial^2 \Phi}{\partial x_i \partial x_j} [f_{\mathrm{K}}(r')\vec{\theta}]. \tag{125}$$

- It is convenient at this point to introduce a weight function

$$W(r, r') \equiv \frac{f_K(r-r') f_K(r')}{f_K(r)}, \tag{126}$$

allowing us to write Equation 125 in the simpler form

$$\frac{\partial \alpha_i}{\partial \theta_j} = \frac{2}{c^2} \int_0^r \mathrm{d}r' W(r, r') \frac{\partial^2 \Phi}{\partial x_i \partial x_j} [f_{\mathrm{K}}(r')\vec{\theta}]. \tag{127}$$

This weight function vanishes at the observer and the source, and peaks approximately half-way between them (or exactly half-way, in the special case of a flat Universe).

- If we take the trace of the matrix defined by Equation 125, then we find that

$$\mathrm{tr} \frac{\partial \alpha_i}{\partial \theta_j} = \frac{2}{c^2} \int_0^r \mathrm{d}r' W(r, r') \frac{\partial^2 \Phi}{\partial x_i^2} [f_{\mathrm{K}}(r')\vec{\theta}] \tag{128}$$

(where we use the usual Einstein convention, so that the sum over i is implied). The two-dimensional Laplacian in this expression can be replaced by its three-dimensional counterpart, since the derivatives along the line-of-sight do not contribute to the integral. Therefore,

$$\mathrm{tr} \frac{\partial \alpha_i}{\partial \theta_j} = \frac{2}{c^2} \int_0^r \mathrm{d}r' W(r, r') \nabla^2 \Phi. \tag{129}$$

- We can now use Poisson's equation to write the $\nabla^2 \Phi$ in terms of the density. Since we are working in comoving coordinates, we need the comoving form of Poisson's equation, which in the limit of small δ is simply

$$\nabla^2 \Phi = 4\pi G \bar{\rho} \delta a^2. \tag{130}$$

We can also write this in terms of H_0 and $\Omega_{\mathrm{m},0}$ as

$$\nabla^2 \Phi = \frac{3}{2} H_0^2 \Omega_{\mathrm{m},0} \frac{\delta}{a}. \tag{131}$$

Equation 129 therefore becomes

$$\mathrm{tr}\frac{\partial\alpha_i}{\partial\theta_j} = \frac{3H_0^2\Omega_{\mathrm{m},0}}{c^2}\int_0^r \mathrm{d}r' W(r,r')\frac{\delta}{a} \equiv 2\kappa, \tag{132}$$

where for reasons that will become clear shortly we have introduced a quantity κ known as the **convergence**.

- The trace-free part of the matrix of deflection angle derivatives can be written as

$$\frac{\partial\alpha_i}{\partial\theta_j} - \frac{1}{2}\delta_{ij}\mathrm{tr}\frac{\partial\alpha_i}{\partial\theta_j} = \frac{\partial\alpha_i}{\partial\theta_j} - \delta_{ij}\kappa \equiv \begin{pmatrix} \gamma_1 & \gamma_2 \\ \gamma_2 & -\gamma_1 \end{pmatrix}, \tag{133}$$

which defines the **shear** components γ_i. Specifically, we have

$$\begin{aligned} \gamma_1 &= \frac{1}{c^2}\int_0^r \mathrm{d}r' W(r,r')\left(\frac{\partial^2\Phi}{\partial x_1^2} - \frac{\partial^2\Phi}{\partial x_2^2}\right), \\ \gamma_2 &= \frac{2}{c^2}\int_0^r \mathrm{d}r' W(r,r')\left(\frac{\partial^2\Phi}{\partial x_1 \partial x_2}\right). \end{aligned} \tag{134}$$

- Combining these results, we see that

$$\frac{\partial\alpha_i}{\partial\theta_j} = \begin{pmatrix} \kappa+\gamma_1 & \gamma_2 \\ \gamma_2 & \kappa-\gamma_1 \end{pmatrix}. \tag{135}$$

- The convergence and the shear correspond to the two different ways in which the properties of the image are affected by gravitational lensing. The convergence corresponds to a change in the size (and hence the brightness) of the image, with the relative magnification being given by

$$\delta\mu = 2\kappa. \tag{136}$$

- The shear component correspond to the image distortion. For example, a circular source with radius R will appear after lensing as an ellipse with major and minor axes

$$a = \frac{R}{1-\kappa-\gamma}, \qquad b = \frac{R}{1-\kappa+\gamma}, \tag{137}$$

where $\gamma \equiv (\gamma_1^2 + \gamma_2^2)^{1/2}$. If the magnification is small, as is usually the case with weak lensing, then the observed ellipticity of a circular source provides a good estimate of the shear:

$$\epsilon \equiv \frac{a-b}{a+b} = \frac{\gamma}{1-\kappa} \simeq \gamma. \tag{138}$$

3.2 Power spectra

- The analysis above demonstrates that measurements of γ and/or κ provide us with information on the second derivatives of the gravitational potential, integrated along the line of sight. Unfortunately, this does not provide us with enough information to allow us to reconstruct either the density field or the potential, since there are many different ways of arranging δ and Φ that would give the same result for the integral along the line of sight. Instead, what weak lensing allows us to do is to constrain the *statistical properties* of the density field.

- To understand how the 3D statistical properties of the density field relate to the 2D properties of the convergence and shear, we need to make use of an important relation known as **Limber's equation**. This states that if we have a density field δ and a projection of that field $g(\vec{\theta})$ that are related by

$$g(\vec{\theta}) = \int_0^r \mathrm{d}r' q(r') \delta[f_{\mathrm{K}}(r')\vec{\theta}], \tag{139}$$

then provided that $q(r)$ changes on scales much larger than δ, the power spectra of g and δ are related by Limber's equation:

$$P_g(l) = \int_0^r \mathrm{d}r' \frac{q^2(r')}{f_{\mathrm{K}}^2(r')} P\left(\frac{l}{f_{\mathrm{K}}(r')}\right). \tag{140}$$

Here, $\mathbf{l}$ is a 2D wave-vector which is the Fourier conjugate to the 2D position on the sky, $\vec{\theta}$.

- Applying this to the case of weak lensing, we find that the power spectrum of the convergence is given by

$$P_\kappa(l) = \frac{9\Omega_{\mathrm{m},0}^2}{4} \frac{H_0^4}{c^4} \int_0^r \mathrm{d}r' \tilde{W}^2(r, r') P\left(\frac{l}{f_{\mathrm{K}}(r')}\right). \tag{141}$$

where

$$\tilde{W}(r, r') \equiv \frac{W(r, r')}{a f_{\mathrm{K}}(r')}. \tag{142}$$

- In practice, we usually do not know the intrinsic luminosities of the lensed galaxies with sufficient precision to be able to determine the magnification and hence the convergence. However, the image distortions cause by shear *can* be measured, albeit only in a statistical sense. Fortunately, it is possible to show that

$$P_\gamma(l) = P_\kappa(l), \tag{143}$$

and so a measure of the shear power spectrum contains information on the power spectrum of the density fluctuations.

3.3 Measurements

3.3.1 Survey requirements

- A fundamental problem that we have to confront if we want to measure γ is that the galaxies that we are observing are not ideal circular sources. In general, they are (to a first approximation) elliptical themselves, and so we need to be able to somehow distinguish between the ellipticity induced by lensing and the intrinsic ellipticity of the galaxies.

- If we write the intrinsic ellipticity as $\epsilon^{(s)}$, then the observed ellipticity for a given galaxy is
$$\epsilon = \epsilon^{(s)} + \gamma. \tag{144}$$
For a single galaxy, there appears to be no practical way to separate $\epsilon^{(s)}$ and γ. However, if we observe a large number of galaxies and compute the mean ellipticity, then we would expect the $\epsilon^{(s)}$ term to average out (to a first approximation), leaving only the mean shear:
$$\langle\epsilon\rangle \simeq \langle\epsilon^{(s)}\rangle + \langle\gamma\rangle \simeq \langle\gamma\rangle. \tag{145}$$

- The typical angular scale on which the cosmic shear signal is strongest is of the order of a few arc-minutes in the ΛCDM model. For our averaging method to work, we therefore need there to be a large number of sources per square arc-minute. Fortunately, deep optical observations can easily recover 30 or more sources per square arc-minute, giving us more than enough to work with.

- A major constraint on the accuracy with which we can measure $\langle\gamma\rangle$ is the fact that because we average only over a finite number of galaxies, $\langle\epsilon^{(s)}\rangle$ is probably not exactly zero. The intrinsic ellipticities of the faint background galaxies have a distribution with standard deviation $\sigma_\epsilon \simeq 0.3$. Therefore, an average over N such galaxies gives a standard error:
$$\delta\epsilon = \langle(\epsilon^{(s)})^2\rangle^{1/2} = \frac{\sigma_\epsilon}{\sqrt{N}}, \tag{146}$$
assuming no correlations exist between the different intrinsic ellipticities.

- We can use this to construct a rough estimate for the signal-to-noise of a cosmic shear measurement. Suppose that we count pairs of galaxies on the sky with separations $\theta \pm \delta\theta$. If the number density of galaxies per unit solid angle is n, then the mean number of galaxies we expect to find with a separation in this range around a randomly selected galaxy is just $2\pi\theta\delta\theta n$, where we have assumed that $\delta\theta \ll \theta$. If our survey covers a total solid angle A, then it will contain a total number of galaxies $N \simeq nA$, and the total number of pairs is then approximately
$$N_p = \frac{1}{2} 2\pi n^2 A\theta\delta\theta, \tag{147}$$
provided that $\theta^2 \ll A$. Therefore, the Poisson noise in our measurement due to the intrinsic ellipticities will be
$$\text{noise} \simeq \frac{2\sigma_\epsilon}{n\sqrt{\pi A\theta\delta\theta}}. \tag{148}$$

- The signal is the square root of the correlation function, which we can approximate as

$$\xi \simeq l^2 P_\kappa(l) \frac{\delta\theta}{\theta}, \tag{149}$$

where $\theta = 2\pi/l$. The signal-to-noise ratio is therefore

$$\frac{S}{N} \simeq \frac{l n \delta\theta \sqrt{\pi A P_\kappa}}{2\sigma_\epsilon} = \frac{n\sqrt{\pi^3 A P_\kappa}}{\sigma_\epsilon} \frac{\delta\theta}{\theta}. \tag{150}$$

- To quote some representative numbers: for $\delta\theta/\theta = 0.1$ and $\theta = 0.1'$, a survey size of one square degree gives a signal-to-noise of order unity. Therefore, to measure the shear to within 10% accuracy, we need a survey of 100+ square degrees, involving tens of millions of galaxies. This is a major undertaking, with the time required for such surveys being of the order of years.

3.3.2 Ellipticity measurements

- So far, we have discussed measuring the ellipticity in fairly abstract terms. However, it is worthwhile looking in more detail at how we actually do this and at some of the problems that can arise.

- In principle, the determination of the ellipticity of a galaxy is a simple task. We measure the surface brightness quadrupole

$$Q_{ij} = \frac{\int I(\mathbf{x}) x_i x_j \mathrm{d}^2 x}{\int I(\mathbf{x}) \mathrm{d}^2 x}, \tag{151}$$

and find its principal axes a and b; the ellipticity then follows from Equation 138.

- In practice, there are a number of problems. First, if the galaxy is small and faint, it may be resolved on our detector by just a few pixels. In this case, the integrals in Equation 151 must be approximated, which inevitably introduces error into our measurement of ϵ. We can avoid this problem by only looking at bright, well-resolved galaxies, but in that case we end up throwing away much of our potential signal, since in any magnitude-limited survey, there are many more faint galaxies than bright galaxies.

- Another issue that must be addressed is the fact that our images of the galaxies are distorted by turbulence in the atmosphere and by imperfections in the telescope optics.

- Atmospheric turbulence has the effect of convolving the actual angular distribution of light from a source with a Gaussian whose width depends on the site of the telescope, the current weather, and many other factors. Typically, the size of this Gaussian – the **seeing** – is around $0.5''$–$1''$. Unless the angular size of the source is much larger than this, the atmospheric distortion has the effect of dramatically reducing the ellipticity.

- A common way of accounting for this effect is to measure and deconvolve the so-called **point-spread function** (PSF) of the telescope, ideally using a measurement of this made at the same time as the actual observations. The PSF describes how the image of a point-like, unresolved source (e.g. a star) appears on the detector. If the telescope optics are even slightly astigmatic, the PSF may be anisotropic, and the degree of anisotropy may depend on the position of the image on the focal plane.

- Systematic errors introduced by these effects can be much larger than the lensing effect we are trying to measure, meaning that any successful measurement must account for and remove them with a high degree of accuracy. This is a highly challenging statistical problem that remains an active area of research.

- Finally, one additional source of spurious shear that we need to understand before we can measure the true shear is a phenomenon known as **intrinsic alignment**. Previously, we assumed that the intrinsic ellipticities of individual galaxies were uncorrelated, meaning that an average over a large number should yield a mean of approximately zero. However, this is not quite correct. Galaxies form within the context of a large-scale gravitational tidal field caused by the surrounding large-scale structure, and this tidal field can lead to nearby galaxies having correlated alignments.

- Considerable work has been devoted to modelling this intrinsic alignment in sufficient detail to allow its effects to be removed from the analysis, and it is no longer seen as a major impediment to measurements of the lensing-induced shear.

4 The cosmic microwave background

4.1 The CMB dipole

- The first anisotropy in the CMB to be detected was the dipole caused by our motion relative to the CMB. If we denote the velocity of the Earth relative to the CMB rest frame by $v_{\oplus}$, then the Doppler shift due to our motion imprints a dipole intensity pattern with an amplitude

$$\frac{\Delta T}{T_0} = \frac{v_{\oplus}}{c}, \tag{152}$$

to first order in $v_{\oplus}/c$.

- The measured amplitude of the dipole is $\simeq 1.24$ mK, from which we can infer that the Earth's velocity relative to the CMB is $v_{\oplus} \simeq 370\,\mathrm{kms}^{-1}$. Once we account for the Earth's motion around the Sun, and the Solar System's motion around the centre of the Milky Way, we find that the Milky Way itself is moving with a higher velocity of $v_{\rm MW} \sim 600\,\mathrm{km\,s}^{-1}$ with respect to the CMB rest frame.

- This motion of the Milky Way is caused by the gravitational attraction of the surrounding matter, which on scales < 100 Mpc is not distributed homogeneously. It therefore tells us something about our local patch of the Universe, but not about the cosmological model itself. To learn about that, we must look at higher order anisotropies in the CMB.

4.2 Primary anisotropies

- We can classify CMB anisotropies into two classes: **primary anisotropies**, which are caused by physical processes acting at the surface of last scattering, and **secondary anisotropies**, which are caused by physical processes acting during the passage of the CMB photons from the last-scattering surface to us.

- We will not attempt in this lecture to give a complete mathematical introduction to the production of either primary or secondary anisotropies. Instead, the aim here is to give a brief pedagogical introduction to the major causes of anisotropies in the CMB. A more detailed mathematical introduction can be found in the review by Hu & Dodelson (2002, ARA&A, 40, 171).

- The basic idea underlying our theory for the production of CMB anisotropies is simple. We assume that the cosmic structures that we see around us today were produced via gravitational instability acting on seed fluctuations in the density field generated in the very early Universe. Since baryons and radiation are strongly coupled prior to recombination, if there are density fluctuations in the baryonic component there must also be density fluctuations in the radiation component.

- Fluctuations in the dark matter component do not coupled directly to the radiation. However, they affect it indirectly, by acting as the sources of fluctuations in the gravitational potential. Therefore, fluctuations in both the baryons and the dark matter lead to fluctuations in the effective temperature of the CMB, i.e. to CMB temperature anisotropies.

- Suppose that there were no dark matter. In that case, a fractional change in the baryon density $\delta\rho/\rho$ should produce a similarly sized change in the radiation energy density. Since the latter scales with temperature as T^4, we therefore have

$$\frac{\delta\rho}{\rho} \simeq \frac{\delta T^4}{T^4} = \frac{4T^3 \delta T}{T^4}. \tag{153}$$

Rearranging this, we see that

$$\frac{\delta T}{T} \simeq \frac{\delta}{4}, \tag{154}$$

where $\delta \equiv \delta\rho/\rho$.

- In an Einstein-de Sitter Universe, density perturbations in the linear regime (i.e. with $\delta < 1$) grow with increasing scale factor as $\delta \propto a$, provided we are in the matter-dominated regime and are considering scales larger enough that we can ignore thermal pressure. A similar relationship holds in a Universe like our own provided that we consider redshifts $z > 1$ where the effects of the cosmological constant are unimportant.

- We therefore expect small density fluctuations at the redshift of recombination to grow by a factor of around 1000 by the present day. Therefore, for these fluctuations to have left the linear regime by the present day, their size at recombination must have been $\delta \geq 10^{-3}$.

- We would therefore expect that in a Universe without dark matter, the size of the temperature anisotropies in the CMB would have been of order 10^{-4} or larger. This is an order of magnitude larger than the values actually observed by COBE and later experiments, and suggests that our simple baryons+radiation model is not compatible with observation.

- Nevertheless, we will persist with this model for a little while longer, as the behaviour of a pure baryons+radiation fluid gives important insight into the more complex behaviour that we see when we have baryons, radiation *and* dark matter.

4.2.1 Acoustic oscillations

- Prior to recombination, the fractional ionization of the gas is $x \sim 1$. At this time, the baryons and the photons are tightly coupled by Thomson scattering. The mean free path (in comoving units) due to Thomson scattering is

$$\lambda_{\rm C} = \frac{1}{n_{\rm e}\sigma_{\rm T} a} \simeq 2.5\,{\rm Mpc}, \tag{155}$$

  which is around a factor of a hundred smaller than the Hubble radius at this time, and so the motion of the photon fluid relative to the baryon fluid is important only on small scales, while on large scales we can assume that the two fluids move with the same velocity.

- To derive a zeroth-order approximation for the behaviour of our coupled baryon-photon fluid, we make three approximations: we assume that the momentum density of the baryons is negligible in comparison to the photons, that the background expansion of the Universe is matter dominated, and that gravitational forces can be neglected.

- These approximations allow us to write down the governing equations for our coupled fluid in a very simple form. First, we know that photons are conserved in Thomson scattering, so the photon number density obeys the continuity equation

$$\dot{n}_\gamma + 3\frac{\dot{a}}{a}n_\gamma + \nabla \cdot (n_\gamma \mathbf{v}_\gamma) = 0, \tag{156}$$

  where the second term is a consequence of the expansion of the Universe. Note also that the dot here represents a derivative with respect to the **conformal time** η, defined as

$$\eta \equiv \int \frac{{\rm d}t}{a(t)}. \tag{157}$$

- If we write the photon number density as the sum of a mean value and a (small) perturbation, $n_\gamma = \bar{n}_\gamma + \delta n_\gamma$, then to first order in our perturbed quantities:
$$\frac{\mathrm{d}}{\mathrm{d}\eta}\left(\frac{\delta n_\gamma}{n_\gamma}\right) = -\nabla \cdot \mathbf{v}_\gamma. \tag{158}$$
Note that in the absence of density perturbations, the velocity $\mathbf{v}_\gamma = 0$ (as otherwise the Universe would not be homogeneous and isotropic).

- For a black-body radiation field, $n_\gamma \propto T^3$, so we can write the photo number density fluctuation in terms of the temperature fluctuation $\Theta \equiv \delta T/T$ as:
$$\frac{\delta n_\gamma}{n_\gamma} = 3\Theta. \tag{159}$$
Therefore, we have
$$\dot{\Theta} = -\frac{1}{3}\nabla \cdot \mathbf{v}_\gamma, \tag{160}$$
which in Fourier space becomes
$$\dot{\hat{\Theta}} = -\frac{1}{3} i\mathbf{k} \cdot \mathbf{v}_\gamma, \tag{161}$$
where $\hat{\Theta}$ is the Fourier transform of Θ and $\mathbf{k}$ is the comoving wavevector.

- The other governing equation for our coupled baryon-photon fluid is the Euler equation. In the absence of gravity, the momentum density of the fluid is altered by only two effects: radiation pressure gradients, which generate the velocity perturbations that accompany our density perturbations, and the expansion of the Universe, which causes the momentum density to decrease with time, owing to the combined effects of the reduction in the photon density and the redshifting of the individual photons. We therefore have
$$\frac{\mathrm{d}}{\mathrm{d}\eta}\left[\left(\rho_\gamma + \frac{p_\gamma}{c^2}\right)\mathbf{v}_\gamma\right] = -4\frac{\dot{a}}{a}\left(\rho_\gamma + \frac{p_\gamma}{c^2}\right)\mathbf{v}_\gamma - \nabla\left(\frac{p_\gamma}{c^2}\right), \tag{162}$$
where $\rho_\gamma c^2$ is the radiation energy density and p_γ is the radiation pressure.

- If we again expand to first order in perturbation theory, we obtain the expression
$$\frac{4}{3}\rho_\gamma \dot{\mathbf{v}}_\gamma = -\frac{1}{3}\nabla \delta\rho_\gamma. \tag{163}$$
where we have used the photon equation of state $p_\gamma = \rho_\gamma c^2/3$ to eliminate p_γ in favour of ρ_γ.

- Since $\rho_\gamma \propto T^4$, it follows that $\delta\rho_\gamma/\rho_\gamma = 4\Theta$, allowing us to write this equation as
$$\dot{\mathbf{v}}_\gamma = -\nabla\Theta, \tag{164}$$
$$\dot{\mathbf{v}}_\gamma = -i\mathbf{k}\hat{\Theta}, \tag{165}$$
$$\tag{166}$$
where the second line gives the Fourier space version of the expression.

- For convenience, we can write $\mathbf{v}_\gamma$ as the product of a scalar velocity and a unit wavevector $\tilde{\mathbf{k}}$, i.e.

$$\mathbf{v}_\gamma \equiv -iv_\gamma \tilde{\mathbf{k}}. \tag{167}$$

This allows us to write the Euler equation as

$$\dot{v}_\gamma = k\hat{\Theta}. \tag{168}$$

- Combining this with our simplified form of the continuity equation then yields

$$\ddot{\hat{\Theta}} + c_{\rm s}^2 k^2 \hat{\Theta} = 0, \tag{169}$$

where $c_{\rm s}^2 \equiv {\rm d}p_\gamma/{\rm d}\rho_\gamma = c^2/3$ is the adiabatic sound-speed of our coupled fluid.

- We therefore see that in this highly simplified model, density perturbations in the coupled fluid undergo simple harmonic oscillation. If we assume that the initial perturbations are adiabatic, with a finite density perturbation but no initial temperature perturbation, then the general solution of Equation 169 can be written as

$$\hat{\Theta}(\eta) = \hat{\Theta}(0)\cos(ks), \tag{170}$$

where

$$s \equiv \int c_{\rm s} {\rm d}\eta \tag{171}$$

is a quantity known as the **sound horizon**, which is simply the comoving distance that a sound wave can propagate in the conformal time η. The simplifying assumptions that we have made above imply that $c_{\rm s}$ does not vary during the period considered, and hence $s = c\eta/\sqrt{3}$.

- In real space, these oscillating Fourier modes correspond to **standing waves**, one for each mode. These oscillations continue until the Universe recombines and the photons and baryons decouple. Following this, the photons interact no further, and hence the state of the oscillation at recombination is encoded in the photon distribution.

- To be slightly more precise, what we observe as inhomogeneities in the CMB are the projection of the 3D spatial fluctuations present at recombination onto a 2D sphere. The mathematical details of this projection are a little involved, but for our purposes, the main results are simply stated. A fluctuation with a physical scale λ_A maps onto an angular scale θ_A according to

$$\theta_A = \frac{\lambda_A}{D_{\rm ang}}, \tag{172}$$

where $D_{\rm ang}$ is the angular diameter distance. Moreover, if we represent our angular fluctuations as a sum of spherical harmonics Y_{lm}, the multipole corresponding to l is sensitive to spatial fluctuations with comoving wavenumbers close to

$$k = \frac{al}{D_{\rm ang}}. \tag{173}$$

- In flat space, the angular diameter distance is directly proportional to the conformal time
$$D_{\rm ang} = a_{\rm rec} c(\eta_0 - \eta_{\rm rec}), \tag{174}$$
where $\eta_{\rm rec}$ is the conformal time at recombination and η_0 is the conformal time at the present day. Since $\eta_0 \gg \eta_{\rm rec}$, this simplifies to $D_{\rm ang} = ac\eta_0$, and hence the comoving wavenumber corresponding to the multipole l is simply
$$k = \frac{l}{c\eta_0}. \tag{175}$$

- From Equation 170, we see that the perturbations with the largest absolute amplitude[3] are those whose comoving wavenumber satisfies:
$$k_n s_{\rm rec} = n\pi, \qquad n = 1, 2, 3... \tag{176}$$
where $s_{\rm rec}$ is the size of the sound horizon at recombination Substituting this into our expression for l yields:
$$l_n = \pi n \frac{c\eta_0}{s_{\rm rec}} \tag{177}$$
Moreover, since we know that $s_{\rm rec} = c\eta_{\rm rec}/\sqrt{3}$, we can rewrite this as
$$l_n = \sqrt{3}\pi \left(\frac{\eta_0}{\eta_{\rm rec}}\right) n. \tag{178}$$

- When dealing with measurements of the CMB anisotropies, it is common to decompose the temperature fluctuations on the sky into a sum of spherical harmonics
$$\Theta(\tilde{\mathbf{n}}) = \sum_{l,m} \Theta_{lm} Y_{lm}(\tilde{\mathbf{n}}), \tag{179}$$
where $\tilde{\mathbf{n}}$ is a unit vector describing the location on the sky, and to then work in terms of the angular power spectrum C_l, defined by
$$\langle \Theta^*_{lm} \Theta_{l'm'} \rangle = \delta_{ll'} \delta_{mm'} C_l, \tag{180}$$
where the angle brackets denote averaging over the whole sky.

- Moreover, one can show that C_l is related to the 3D dimensionless power spectrum $\Delta^2_T(k)$ via
$$\frac{l(l+1)}{2\pi} C_l \simeq \Delta^2_T(k), \tag{181}$$
where $k = l/D_{\rm ang}$. Consequently, when you see the CMB angular power spectrum plotted, what is often shown is $l(l+1)C_l/2\pi$, rather than just C_l.

[3] The power in an observed mode scales as the scale of the amplitude, so both maxima and minima contribute

- The key result that our simple model of the baryon-photon fluid has lead us to is the following: *oscillations in the baryon-photon fluid will produce a series of peaks and troughs in the power spectrum of the fluctuations, with the spacing between the peaks determined primarily by the ratio of η_0 to $\eta_{\rm rec}$.*

- In our simplified model, we have assumed that the expansion of the Universe is matter-dominated throughout. In this case, $\eta \propto (1+z)^{-1/2}$, and so $\eta_0/\eta_{\rm rec} = (1+z_{\rm rec})^{1/2} \simeq 33$. This leads to the first **acoustic peak** occurring at $l_n \simeq 180$.

- A more careful calculation that accounts for the effects of radiation near $z_{\rm rec}$, the effects of the cosmological constant at low z, and that does not assume a flat Universe demonstrates that the exact position of the peak is sensitive to the values of several of the cosmological parameters. Most importantly, it is sensitive to the value of $\Omega_{\rm tot} \equiv \Omega_{\rm m} + \Omega_{\Lambda}$, as changing this can lead to a considerable change in the angular diameter distance.

- We can therefore use measurements of the position of the first peak in the CMB power spectrum to constrain the geometry of the Universe. The first detections of this peak, by the balloon-borne experiments BOOMERANG and MAXIMA, had insufficient precision to provide a strong constraint on $\Omega_{\rm tot}$, but the subsequent measurements by WMAP and PLANCK do provide strong constraints, and show that $\Omega_{\rm tot} \simeq 1$ to within a precision of a few percent; i.e. the Universe is very close to being flat.

4.2.2 The Sachs-Wolfe effect

- We next explore what happens if we relax some of the simplifying assumptions made previously. We start by examining what happens if we include gravitational effects.

- Suppose that we slightly perturb the otherwise flat gravitational potential. GR tells us that our perturbed potential Ψ acts as a perturbation of the time coordinate:

$$\frac{\delta t}{t} = \Psi. \tag{182}$$

- This perturbation has two important effects. First, it acts as a perturbation in the scale factor. In the matter-dominated era, we have:

$$\frac{\delta a}{a} = \frac{2}{3}\frac{\delta t}{t}, \tag{183}$$

  since $a \propto t^{2/3}$ in this era. Moreover, since the CMB temperature is related to the scale factor as $T \propto a^{-1}$, the perturbation in the potential leads to a local temperature perturbation

$$\Theta = -\frac{2}{3}\Psi. \tag{184}$$

  In other words, in regions where Ψ is negative (potential wells), we have a local overdensity of photons and hence a local increase in the temperature. Similarly, in underdense regions (with positive Ψ), we have a temperature deficit.

- We therefore have two equivalent ways to think about the same set of perturbations: we can think of them as density perturbations that generate corresponding curvature perturbations; or we can think of them as curvature perturbations that generate density perturbations through their effect on the scale factor.

- However, aside from its effects on the local density, the curvature also affects the observed temperature perturbations in another way. Photons in an overdense, hotter region are located in a potential well compared to their surroundings, and must climb out of this potential well before we can observe them. As they do so, they undergo a **gravitational redshift** leading to a second temperature perturbation of magnitude Ψ.

- The temperature perturbation that we observe is simply the sum of the two terms, one due to the photon overdensity and a second due to the gravitational redshift. We therefore see that:
$$\Theta = \frac{1}{3}\Psi. \tag{185}$$

- Since the gravitational redshift has a larger effect than the photon overdensity, we see that hot spots in the observed pattern of temperature anisotropies actually correspond to underdense regions, while cold spots correspond to overdense regions.

- It is possible to show (although we will not do so here) that the time evolution of Ψ satisfies the same oscillator equation as Θ did in our simplified model, i.e.
$$\ddot{\hat{\Psi}} + c_{\rm s}^2 k^2 \hat{\Psi} = 0, \tag{186}$$
where $\hat{\Psi}$ is the Fourier component of Ψ corresponding to wavenumber k. Therefore, we have the same formal solution for $\hat{\Psi}$ as we did for $\hat{\Theta}$, namely
$$\hat{\Psi}(\eta) = \hat{\Psi}(0)\cos(ks), \tag{187}$$
and the same phenomenology, i.e. the same series of acoustic peaks.

- On very large scales, k is small and $\cos(ks) \sim 1$. In this limit, we therefore see that $\hat{\Psi}(\eta) \simeq \hat{\Psi}(0)$, and hence that the perturbations that we observe are directly probing the initial curvature fluctuations. For historical reasons, we often talk about these large-scale perturbations as being generated by the **Sachs-Wolfe** effect, while ascribing the smaller-scale perturbations to the effects of **acoustic oscillations**, but as we have seen in this treatment, the physical processes are the same in both cases, and so this is a rather arbitrary distinction.

4.2.3 Baryon loading

- The next improvement we can make to our treatment is to account for the effects of the baryon momentum. Baryons enhance the momentum density of the coupled fluid by a factor
$$R = \frac{p_b + \rho_b c^2}{p_\gamma + \rho_\gamma c^2}. \tag{188}$$

If we assume that the baryons are non-relativistic, so that $p_b \simeq 0$, then we can write this in terms of $\Omega_{\mathrm{b},0}$ and $\Omega_{\mathrm{r},0}$ as

$$R = \frac{3}{4(1+z)} \left(\frac{\Omega_{\mathrm{b},0}}{\Omega_{\mathrm{r},0}} \right). \tag{189}$$

Substituting in approximate values for $\Omega_{\mathrm{b},0}$ and $\Omega_{\mathrm{r},0}$, we find that $R \sim 700/(1+z)$, and hence is of order unity at recombination.

- Including the baryons has no effect on the photon density continuity equation, but does change the Euler equation, which becomes

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\eta} \left[(1+R) \left(\rho_\gamma + \frac{p_\gamma}{c^2} \right) \mathbf{v}_{\gamma,b} \right] = &- 4\frac{\dot{a}}{a}(1+R) \left(\rho_\gamma + \frac{p_\gamma}{c^2} \right) \mathbf{v}_{\gamma,b} - \nabla \left(\frac{p_\gamma}{c^2} \right) \\ &- (1+R) \left(\rho_\gamma + \frac{p_\gamma}{c^2} \right) \nabla \Psi, \end{aligned} \tag{190}$$

  where $\mathbf{v}_{\gamma,b}$ is the momentum-weighted velocity of the coupled fluid.

- Using this Equation and the continuity equation, It is possible to show that the perturbations still satisfy an oscillator equation. However, in the presence of baryons, the sound speed of the oscillations is slightly different,

$$c_{\mathrm{s}}^2 = \frac{c^2}{3(1+R)}, \tag{191}$$

  meaning that the first acoustic peak is shifted to higher l by a factor $\sqrt{1+R}$, corresponding to around a 30% change.

- In addition, the additional gravitational attraction of the baryons deepens the potential wells and leads to greater gravitational redshifting of the photons. The effect is to increase the height of the acoustic peaks that correspond to compressions (i.e. the odd-numbered peaks) and to decrease the height of the even-numbered peaks.

- We can therefore use the relative height of the acoustic peaks in the CMB power spectrum to constrain R and hence $\Omega_{\mathrm{b},0}$ (since $\Omega_{\mathrm{r},0}$ is already well constrained). The value for $\Omega_{\mathrm{b},0}$ that we obtain in this way agrees very well with the value that we obtain from nucleosynthesis, providing further evidence that our basic cosmological picture is consistent.

4.2.4 Doppler effect

- Since the baryon-photon fluid is moving relative to the observer, we might expect to see temperature fluctuations due to an additional effect, the **Doppler effect**. Regions where the fluid is moving away from us should give rise to redshifted photons and hence lower temperatures, while regions where the plasma is moving towards us will give rise to blueshifted photons and hence higher temperatures.

- A simple estimate of the size of this effect suggests that it should be rather large. We know that in our simplified model, $v_\gamma \propto \dot{\hat{\Theta}}/k$, and we know also from our solution for $\hat{\Theta}$ that $\dot{\hat{\Theta}} \propto kc\hat{\Theta}$. Finally, we expect the size of the temperature perturbations that are generated by the Doppler effect to be proportional to v_γ/c.

- Putting this all together, we see that the size of the temperature perturbations due to the Doppler effect is in principle comparable to the size of the temperature perturbations generated by the overdensity itself.

- However, there is one important difference in the behaviour of the Doppler effect term compared to that of the overdensity term. With the Doppler effect, we only see a significant signal when we are looking in a direction that is close to that of the velocity: motion in the plane of the sky has no effect on the signal we see, only motion perpendicular to that plane.

- In addition, if we think about the behaviour of our fluid as it falls into a potential well, it is easy to see that we will generally have motions both towards and away from us, whose effects on Θ will largely cancel.

- This directional dependence leads to the Doppler effect contribution to the anisotropies and the observed power spectrum being strongly smoothed out in k-space. The observed pattern of peaks is therefore still dominated by the acoustic effects discussed above, and the Doppler term merely contributes some additional, smoothly-distributed power.

4.2.5 Damping

- One final important effect that we need to include in our model is **damping**. So far, we have treated the baryons and the photons as a perfectly coupled fluid. On large scales, this is a reasonable approximation, but this description breaks down on scales comparable to the Thomson scattering mean free path.

- On small scales, CMB photons undergoing repeated scatterings execute a random walk. After N scatterings, the average photon has diffused a distance

$$\lambda_{\rm D} = \sqrt{N}\lambda_{\rm C}, \tag{192}$$

where $\lambda_{\rm C}$ is the Thomson scattering mean free path. Since the mean number of scatterings that a given photon undergoes is

$$N = \frac{c\eta}{\lambda_{\rm C}}, \tag{193}$$

we see that

$$\lambda_{\rm D} = \sqrt{c\eta\lambda_{\rm C}}. \tag{194}$$

As a fraction of the sound horizon, this is:

$$\frac{\lambda_{\rm D}}{c\eta_{\rm rec}} = \sqrt{\frac{3\lambda_{\rm C}}{c\eta_{\rm rec}}}, \tag{195}$$

which is around a few percent. We therefore expect features in the power spectrum at $l \sim 10^3$ and above to be strongly affected by damping.

- The effect of this damping, often referred to as **Silk damping**, is to exponentially suppress features in the power spectrum at large l, corresponding to the acoustic peaks with $n > 3$.

4.3 Secondary anisotropies

- The primary anisotropies discussed in the previous section are all generated prior to cosmological recombination. At this time, the fluctuations that we are dealing with are small, and can be treated using the tools of linear perturbation theory. Therefore, even though a detailed mathematical treatment can become rather complex, we can in principle predict the pattern of anisotropies to within a very high degree of precision.

- The same cannot be said for the secondary anisotropies. These are generated by physical processes occurring after recombination, and therefore often probe the non-linear regime of structure formation, which we cannot model so accurately.

- There are two main ways in which secondary anisotropies can be introduced into the CMB. The first of these involves gravitational redshifting of the CMB photons. Suppose that a photon propagating to us passes through a deep potential well. As the photon moves down into the potential well, it will be blueshifted, while as it climbs out again, it will be redshifted. If the potential well is static, then there is no net effect: the blueshifting and redshifting exactly cancel. However, if the potential is varying with time, then the amount of blueshifting and redshifting need not exactly agree, leading to a net change in the energy of the photon.

- Therefore, changes with time in the gravitational field through which the CMB photons are propagating lead to new anisotropies in the photon temperature distribution. While the overdensities and underdensities responsible for this effect remain in the linear regime, we refer to is as the **integrated Sachs-Wolfe effect**. On the other hand, when we are dealing with highly non-linear structures (e.g. galaxy clusters), we refer to the same effect as the **Rees-Sciama effect**.

- We can quantify both forms of the effect with the help of large N-body simulations. These show that at low l, the contribution made by the ISW or RS effects is orders of magnitude smaller than the contribution made by the primary anisotropies. However, at large l the primary anisotropies are exponentially damped, while the ISW and RS effects are not. We therefore expect them to become important for $l \sim 5000$ and above.

- The other main source of secondary anisotropies is Compton scattering of the CMB photons by rapidly moving electrons. Although most CMB photons scatter for the last time at the surface of last scattering, the probability of subsequent scattering is not completely zero, and so a small fraction of photons scatter at later times.

- If the scattering is elastic (i.e. Thomson scattering), then the net effect is to partially damp the primary anisotropies, by mixing together photons from hot regions and cold regions. However, if the scattering is inelastic (i.e. Compton scattering), so that the CMB photons either gain or lose energy, then new anisotropies can be created.

- The general name for this process is the **Sunyaev-Zel'dovich effect**, or SZ effect for short, and it comes in two varieties. To create significant anisotropies, we need rapidly moving electrons. If the electron velocities are high because the gas is very hot, then we have the **thermal SZ effect**; on the other hand, if the velocities are high because of some bulk flow of the ionized gas, then we have the **kinetic SZ effect**.

- In general, the only places where we find sufficient amounts of hot gas to create detectable anisotropies via the thermal SZ effect are galaxy clusters. The angular size of these clusters on the sky is small, so the thermal SZ effect only plays a role at very high l. Moreover, given a detector with sufficiently high angular resolution, individual clusters can be detected, allowing them to be masked out of the larger-scale maps.

- The kinetic SZ effect is generated on larger scales, by the infall of ionized gas into overdense regions such as clusters or superclusters of galaxies. It's importance can be quantified using numerical simulations. As with the RS effect, the kinetic SZ effect is unimportant compared to the primary anisotropies at values of l less than a few thousand, but can become significant at very high l.

4.4 Polarization

- For any given patch of the CMB, we can measure not only its intensity, but also its polarization. In a perfectly homogeneous and isotropic Universe, the mean polarization everywhere would be zero, i.e. we would expect to find equal numbers of photons with both different polarization directions.

- In an inhomogeneous Universe, this is no longer true, for a reason that we will now explain. We start with the fact that Thomson scattering is polarization sensitive, and can produce polarized radiation from unpolarized radiation. Scattering of an unpolarized beam of light into our line of sight produces a polarized signal, with the details of the net polarization depending on the direction of the incoming beam relative to our line of sight. If the electrons responsible for the scattering are uniformly illuminated, then the effect cancels out. However, if the incoming radiation is anisotropic, then the net effect is no longer zero.

- In the case of the CMB, we therefore expect a polarization signal whose strength is sensitive to the size of the anisotropies in the CMB. Detailed calculations show that the polarized signal should have a strength which is roughly 10% of that of the unpolarized anisotropies, i.e. around 10^{-6}.

- The tiny size of this polarization makes it difficult to detect, but a number of recent CMB experiments have been able to detect it, providing further confirmation of our picture of the CMB's origin. PLANCK has the sensitivity to revolutionize this topic, but has not yet released any results regarding the polarization of the CMB, as the PLANCK collaboration is still engaged in fully understanding all of the possible sources of systematic error in the measurement.

- In addition to this intrinsic polarization, there is an additional source of polarization that is somewhat easier to detect. Once the Universe becomes reionized, at a redshift $z \sim 10$, there is a small probability that CMB photons will scatter off the electrons in the IGM. The optical depth due to this scattering is relatively small, $\tau \sim 0.1$, but is large enough to affect the CMB. It acts to damp the anisotropies, but on its own, this is difficult to distinguish from a change in the initial amplitude of the perturbations. However, it also introduces a net polarization, for the same reason as above. Detecting this polarization therefore allows one to constrain τ and hence the redshift of reionization.

4.5 Foregrounds

- It took 20+ years of increasingly sensitive measurements to go from a detection of the CMB itself to a detection of CMB anisotropies beyond the dipole. It then took another 10+ years until we could first start to map out the CMB power spectrum in any detail, first with balloon-borne experiments such as BOOMERANG or MAXIMA, and then with the launch of the WMAP satellite.

- Why did it take so long? The extremely small size of the anisotropies means that in order to measure them, our instruments need to be extremely well calibrated. Relative temperature errors between one part of the sky and another must be made smaller than one part in 10^5. In addition, **foreground contamination** must be removed with a similar level of accuracy.

- What are possible sources of foreground contamination? Basically, any present-day source that emits microwave photons, or any higher redshift source that emits photons that by now have red-shifted into the microwave regime.

- Foreground sources can be grouped into two main types: **point sources** and **diffuse sources**. Important examples of point sources include high-redshift infrared-bright galaxies, galaxy clusters (which affect the CMB through the Sunyaev-Zel'dovich effect) and minor planets in our own Solar System. Important examples of diffuse sources include synchrotron emission from relativistic electrons, bremsstrahlung from ionized gas, and thermal dust emission.

- Point sources are relatively easy to deal with. The sources responsible (galaxy clusters, high-z galaxies etc.) can be identified in separate surveys, allowing one to identify which parts of the sky have point source contamination. These regions can then simply be excluded from the subsequent analysis.

- Diffuse emission is more of a challenge to deal with. There are two main strategies that are used to mitigate their effect. First, we typically observe CMB anisotropies at frequencies where the foregrounds are least important. This works because the frequency dependence of the foregrounds is generally very different from that of the CMB. For example, thermal dust emission can be represented with a modified black-body spectrum
$$I_\nu = \tau_{\nu_0} \left(\frac{\nu}{\nu_0}\right)^\beta B_\nu(\nu, T_{\rm d}), \tag{196}$$
where τ_{ν_0} is the dust optical depth at the reference frequency ν_0. Typically, Galactic dust has a temperature of around $T_{\rm d} \sim 20{\rm K}$, and so dust emission peaks at a wavelength of around 200 μm, significantly shorter than the peak wavelength of the CMB spectrum, which is located at $\lambda \sim 2$ mm.

- At long wavelengths, synchrotron emission from relativistic electrons interacting with the Galactic magnetic field dominates the microwave sky. This has a power-law spectrum that behaves approximately as $I_\nu \propto \nu^{-1.5}$, and hence becomes much less important as we move to shorter wavelengths.

- Fortunately, between these two sources of emission, there is a relatively clear window, ranging from 30–200 GHz, and for this reason studies of CMB anisotropies generally try to focus on this window. Nevertheless, some foreground contamination remains even within this window. To deal with this contamination, the usual strategy is to observe the CMB at a range of different frequencies. Since the frequency-dependence of the thermal dust emission and the synchrotron emission both differ from that of the CMB emission, by producing appropriately weighted sums and differences of our different measurements, we can remove the contamination, leaving only the CMB signal.

- Finally, an important point to bear in mind is that although foregrounds are generally "noise" as far as cosmologists are concerned, they actually contain considerable astrophysical information and can be of great interest in their own right. For example, the PLANCK CMB maps have been used to construct all-sky maps of CO emission in the lowest three rotational emission lines, which are of great interest to those of us studying the Galactic ISM.

5 Galaxy clustering

5.1 Motivation

- The CMB gives us information on density perturbations while they're still in linear regime, and hence (relatively) easy to treat However, small-scale modes in the density power spectrum are strongly affected by damping due to electron scattering in foreground, other sources of foreground noise Therefore, the CMB primarily gives us information on the largest scale modes, down to scales of the order of 10^{-3} times the Hubble radius.

- These large-scale modes correspond to spatial scales of order 10 Mpc at the present day. Studying clustering in the galaxy distribution on these scales therefore provides a useful check on the CMB results.
- In addition, the growth of clustering depends on the growth factor D_+, which is sensitive to Ω_Λ, while the CMB gives no *direct* information on Λ, since its contribution is negligible at $z \sim 1100$.
- Furthermore, we can probe smaller scales with galaxy clustering than with CMB, allowing us to test the CDM model over a wider range of scales.

5.2 Two-point correlation function

- We can define a correlation function for the density field as

$$\xi(y) \equiv \langle \delta(\vec{x})\delta(\vec{x}+\vec{y}) \rangle \tag{197}$$

where the average extends over all spatial positions $\vec{x}$ and orientations of $\vec{y}$; because of isotropy, ξ depends only on the magnitude of $\vec{y}$, and not its orientation.

- By writing $\delta(\vec{x})$ and $\delta(\vec{x}+\vec{y})$ in terms of their Fourier transforms, we can derive an expression for ξ in terms of an integral over the power spectrum $P(k)$:

$$\xi(y) = 4\pi \int \frac{k^2 \mathrm{d}k}{(2\pi)^3} P(k) \frac{\sin ky}{ky}. \tag{198}$$

In other words, $\xi(y)$ is completely determined once $P(k)$ is known.

- However, this is the correlation function of the *density field*, which is not directly observable – if dark matter dominates, we can infer it only from its effects on the luminous matter. We therefore want to know how to relate the correlation function of the dark matter to the correlation function of something that we can observe, such as the number density of galaxies or galaxy clusters.
- The form of ξ written above assumes that δ is a continuous field, which is reasonable when dealing with the density field. On the other hand, galaxies are clearly individual objects – it doesn't make sense in this context to talk about half a galaxy. So how do we define a correlation function for an inherently integral object such as a galaxy?
- *Answer*: work in terms of probabilities, which can still be continuous. Define the **two-point correlation function** $\xi(r)$ to be the excess probability for finding a neighbour

at a distance r from a given galaxy. In other words, we write the probability for finding a pair of objects in the small volume elements dV_1 and dV_2, separated by r, as

$$dP = \bar{n}_{\rm gal}^2[1 + \xi(r)]dV_1 dV_2, \tag{199}$$

where $\bar{n}_{\rm gal}$ is the mean number density of galaxies

- If galaxies are randomly distributed, then the probability of finding a galaxy in the volume element dV_1 is
$$dP_1 = \bar{n}_{\rm gal} dV_1, \tag{200}$$
the analogous expression for volume element dV_2 is
$$dP_2 = \bar{n}_{\rm gal} dV_2, \tag{201}$$
and hence
$$\begin{aligned} dP &= dP_1 dP_2 && (202)\\ &= \bar{n}_{\rm gal}^2 dV_1 dV_2, && (203)\end{aligned}$$
demonstrating that $\xi(r)$ is zero for a random distribution.

- If we have a number density of galaxies that is directly proportional to the density field, i.e.
$$n_{\rm gal}(\vec{x}) = A\rho(\vec{x}), \tag{204}$$
where A is a constant of proportionality, then we can write dP as
$$dP = A^2 \left\langle \rho(\vec{x})\rho(\vec{x} + \vec{r}) dV_1 dV_2 \right\rangle, \tag{205}$$
where the average is over all positions $\vec{x}$ and orientations of $\vec{r}$. We can use the fact that $\rho(\vec{x}) = [1 + \delta(\vec{x})]\, \rho_{\rm b}$, where ρ_b is the background density, to write this expression as
$$\begin{aligned} dP &= A^2 \rho_b^2 \left\langle [1 + \delta(\vec{x})]\, [1 + \delta(\vec{x} + \vec{r})] \right\rangle dV_1 dV_2 && (206)\\ &= \bar{n}_{\rm gal}^2 \left\langle (1 + \delta(\vec{x}) + \delta(\vec{x} + \vec{r}) + \delta(\vec{x})\delta(\vec{x} + \vec{r}) \right\rangle dV_1 dV_2. && (207)\end{aligned}$$
Mass conservation implies that the spatial average of the density contrast, $\langle \delta \rangle$, must vanish, and so we can reduce this to:
$$dP = \bar{n}_{\rm gal}^2 \left[1 + \langle \delta(\vec{x})\delta(\vec{x} + \vec{r}) \rangle\right] dV_1 dV_2 \tag{208}$$
Now, $\langle \delta(\vec{x})\delta(\vec{x} + \vec{r}) \rangle$ is just the correlation function of the density field, and so we see that in this simple case, where the number density of galaxies just traces the underlying density field, the two-point correlation function of the galaxies on a scale r is just the same as the two-point function of the density field.

- In general, however, the number density of galaxies is *not* directly proportional to the underlying density field – galaxies are a *biased* tracer of the density.

- Before going on to look at this idea in more detail, we should also note that we can define a function that is closely related to the two-point correlation function that is useful when we want to compare different *classes* of objects
- We define the **cross-correlation function** ξ_{ab} of two different classes of objects a and b as
$$dP = \bar{n}_a \bar{n}_b \left[1 + \xi_{ab}(r)\right] dV_1 dV_2, \tag{209}$$
where $\bar{n}_a$ and $\bar{n}_b$ are the mean number densities of objects from class a and class b, respectively. We can use this cross-correlation function to quantify e.g. the tendency to find more galaxies in regions close to galaxy clusters than in the field.

5.3 Galaxies as peaks in the density field

- Galaxies form inside dark matter halos, which themselves represent regions in the underlying density field in which the evolution of the density perturbations has become strongly non-linear
- In fact, dark matter halo represent *peaks*, i.e. local maxima, of the underlying density field
- If the underlying density field is a Gaussian random field, then we can solve for various interesting quantities, such as the mean number density of local maxima.
- The maths involved is rather time-consuming (see Bardeen et al. 1986 for the full details), but the end result is that for high peaks (i.e. peaks significantly higher than the density variance), the number density of peaks with overdensities greater than $\nu \equiv \delta/\sigma$ can be written as
$$n_{\rm pk}(> \nu) \propto \nu^2 \exp\left(-\frac{1}{2}\nu^2\right) \tag{210}$$
Note the similarity of this expression to the Press-Schechter mass function
$$N(M)dM = \sqrt{\frac{2}{\pi}} \frac{\rho \delta_{\rm c}}{\sigma} \left|\frac{{\rm d}\ln\sigma}{{\rm d}M}\right| \exp\left(-\frac{\delta_{\rm c}^2}{2\sigma^2}\right) \frac{{\rm d}M}{M}, \tag{211}$$
which we can write in the form
$$N(M) \propto \nu_{\rm c} \exp\left(-\frac{1}{2}\nu_{\rm c}^2\right), \tag{212}$$
where $\nu_{\rm c} = \delta_{\rm c}/\sigma$.
- If galaxies and clusters correspond to peaks in the underlying density field, then this implies that they will be inherently *more* clustered than the underlying density.

- A simple picture that allows us to see this is known as the **peak-background split**. We conceptually decompose the density field into short wavelength terms, corresponding to the peaks plus terms of much longer wavelength, which modulate the peak number density.
- If there is some threshold density ν for the formation of galaxies, then this will be easier to reach in regions where the long wavelength modes contribute positively rather than negatively.
- It is unclear *a priori* exactly what threshold is appropriate, but if galaxies/clusters are to be identified with peaks, then this implies that $\nu_{\rm th} \geq 1$.
- Hence, even if galaxies & clusters trace underlying mass distribution, we expect them to be *more* clustered than the mass.
- We can use the peak-background split to estimate the effect on the correlation function, provided we're dealing with scales where linear evolution can still be assumed (e.g. clustering of galaxy clusters; clustering of galaxies strongly affected by non-linear effects).
- In regions where long wavelength modes contribute positively, we can treat this as a lowering of the effective threshold for collapse:

$$\nu_{\rm c,eff} = \nu_{\rm c} - \frac{\delta_{\rm bk}}{\sigma} \tag{213}$$

  where $\delta_{\rm bk}$ is the contribution from the long wavelength modes (the "background").
- To compute effect of this on the halo number density, consider the number of halos of a given mass M to be a function of the collapse threshold $\nu_{\rm c}$, and Taylor expand to first order:

$$N_M(\nu_{\rm c,eff}) = N_M(\nu_{\rm c}) - \frac{\delta_{\rm bk}}{\sigma}\frac{{\rm d}N_M}{{\rm d}\nu} \tag{214}$$

$$= N_M(\nu_{\rm c})\left[1 - \frac{\delta_{\rm bk}}{\sigma}\frac{1}{N_M(\nu_{\rm c})}\frac{{\rm d}N_M}{{\rm d}\nu}\right]. \tag{215}$$

  In the high peak limit, we have ${\rm d}\ln N_M(\nu_{\rm c})/d\nu \simeq -\nu_{\rm c}$, and hence the equation above becomes

$$N_M(\nu_{\rm c,eff}) = N_M(\nu_{\rm c})\left[1 + \frac{\nu_{\rm c}\delta_{\rm bk}}{\sigma}\right]. \tag{216}$$

  This is equivalent to a perturbation in the number density of the form

$$\frac{\delta n}{n} \simeq \frac{\nu_{\rm c}\delta_{\rm bk}}{\sigma}, \tag{217}$$

  and since the two-point correlation function scales as the square of this, we finally conclude that the correlation function of the peaks can be written in terms of the correlation function of the mass as

$$\xi_{\rm peak}(r) \simeq \frac{\nu_{\rm c}^2}{\sigma^2}\xi_{\rm mass}(r) \tag{218}$$

[NB. This assumes we're considering scales r that are large enough that we're still in the linear regime].

- More generally, the correlation function of a peak with height ν (which need not be ν_c) is given, in the high-peak limit, by

$$\xi_{\rm peak}(r) \simeq \frac{\nu^2}{\sigma^2}\xi_{\rm mass}(r) \tag{219}$$

- This difference between the correlation function of the peaks and the background is often referred to as **bias**. If galaxies form only in the peaks, they are then a *biased* tracer of the mass distribution This is often quantified in terms of a linear bias parameter b, defined such that

$$\xi_{\rm gal} = b^2\xi_{\rm mass} \tag{220}$$

Note that b need not be the same for all classes of objects!

5.4 Non-linear clustering

- On the very largest scales, the distribution of the richest galaxy clusters can still be well-described by linear theory.
- On smaller scales, linear theory breaks down and evolution of density field becomes highly non-linear. Nevertheless, there are various techniques we can use to model the evolution of $\xi(r)$ in the non-linear regime
- Consider a simple toy model for the distribution of galaxies: we place all galaxies in collapsed, virialized clusters, and treat the clustering of these clusters using linear theory, as above
- In this model, correlation function on small scales is dominated by signal coming from clusters. If clusters have very steep, power-law profiles $\rho \propto r^{-\eta}$, then most galaxies lie near the center of clusters and we can write the correlation function as $\xi(r) \propto r^{-\gamma}$, where $\gamma = \eta$. [To see this, consider a galaxy lying at the center of the cluster. The probability that there is a second galaxy at a distance r is proportional to the number density of galaxies at that r, i.e. it is proportional to $\rho(r)$. The correlation function *for that galaxy* therefore falls off as $\rho(r)$. But if most of the galaxies are in the cluster centre, then the same is true for most galaxies, and hence the overall correlation function just scales as the density].
- In the more general case, it is possible to show that the relationship between γ and η is given by $\gamma = 2\eta - 3$ for η in the range $3/2 < \eta < 3$.
- In any case, critical point is that small-scale behaviour of ξ is determined directly by the density profile of the larger-scale objects (in this case, the galaxy clusters)

- How does ξ evolve with time in this model? The hypothesis of **stable clustering** states that although the separation of clusters will change as the universe expands, their internal structure will not (i.e. in this approximation, we ignore the fact that the clusters may be growing with time).

- In this approximation, overdensity of cluster with respect to background evolves as $(1+z)^3$, but cluster size fixed in *proper* coordinate, meaning that if we consider correlation function at some fixed *comoving* coordinate, we will probe different length scales within the cluster.

- Putting these two effects together, and assuming that the *proper* correlation function within the clusters scales as $\xi(r) \propto r^{-\gamma}$, we find that the *comoving* correlation function evolves with redshift as
$$\xi(x, z) \propto (1 + z)^{\gamma-3} \tag{221}$$

- In the linear regime, we have instead
$$\xi(x, z) \propto D_+^2 (1 + z)^{-2} \tag{222}$$

- The linear and non-linear regimes must match at the **scale of quasi-linearity**, r_0, defined by $\xi(r_0) = 1$, and must agree on how this scale evolves with redshift.

- Linear theory gives $r_0 \propto (1 + z)^{-2/(n_{\rm eff}+3)}$, where $n_{\rm eff}$ is the effective index of the power spectrum on the relevant scales, while stable clustering implies $r_0 \propto (1+z)^{-(3-\gamma)/\gamma}$. By equating the indices, we can solve for the index γ of the non-linear correlation function, obtaining
$$\gamma = \frac{3n_{\rm eff} + 9}{n_{\rm eff} + 5}. \tag{223}$$

- Observations suggest that $\gamma = 1.8$, implying that on the scale of galaxy clusters, $n_{\rm eff} \sim 0$.

5.5 Halo model

- We can generalize the ideas outlined above, and construct a model known as the halo model. The basic idea is very simple: we assume that the small-scale clustering of the dark matter is determined by the density profile *within* halos, and that the large-scale clustering is determined by the clustering *of* halos.

- However, not every halo will contain the same number of galaxies; not every halo has same mass. We therefore introduce a quantity $P(N|M)$ that gives the probability that a halo of mass M holds a number of galaxies N.

- If we combine this with a model for the spatial and velocity distribution of the galaxies within the halo, then we can solve for whatever clustering statistics we want. In practice, these quantities are usually constrained by numerical simulations.

- Typically, we assume that $P(N|M)$ is solely a function of M – i.e. the number of galaxies living in a given halo is just a function of the size of the halo. This is probably not true for any individual halo, but we can hope that it's true in an average sense (i.e. that history effects average out).
- Testing the halo model is an important area of research in cosmology right now.

5.6 Redshift-space effects

- Even if we can relate the correlation function of galaxies to the correlation function of the underlying matter distribution, we're then immediately faced with another problem: *we cannot measure* $\xi_{\rm gal}(r)$.
- The correlation function is defined in terms of an average over space, but we don't know the true spatial positions of the galaxies we observe. Instead, the positions of the galaxies are defined in terms of two angular coordinates (the angular position on our sky) and a *redshift*.
- To convert from redshift to radial distance (and hence spatial position), we use Hubble's law. However, this does not give us exact positions, because of the distorting effects of the peculiar velocities of the galaxies.
- Consider a galaxy at a radial distance r. According to Hubble's law, it has a velocity $v = Hr$ (in the low-redshift limit), and an associated redshift which we shall call z_H. If this galaxy has a peculiar velocity $v_{\rm pec}$ along our line of sight to it, then the redshift we actually measure will be given by

$$1 + z_{\rm obs} = (1 + z_{\rm H}) \left(1 + \frac{v_{\rm pec}}{c}\right) \tag{224}$$

- If the peculiar velocities were randomly distributed, this would simply add a noise term to our redshift determinations, and we could deal with the problem simply by making our galaxy sample sufficiently large
- However, in practice this doesn't work, because the distribution of peculiar velocities is not random. As we have already seen, density and velocity perturbations are correlated, and so the effect of the peculiar velocities is to create a systematic distortion of the clustering pattern observed in **redshift space** compared to the pattern in **real space**.
- Fortunately, the close relationship between density and velocity perturbations makes it easy to account for the effects of the latter in our analysis, at least in the linear regime.
- Suppose that we're dealing with a distant region of space, so that the small angle approximation holds, and our radial distortions can be considered as occurring along a single Cartesian coordinate axis. Then we can use the Zel'dovich approximation to write the peculiar velocity as

$$\vec{u} = Hf(\Omega_{\rm m})\vec{f}, \tag{225}$$

where $\vec{f}$ is the displacement field and $f(\Omega_{\rm m}) \simeq \Omega_{\rm m}^{0.6}$.

- Using the Zel'dovich approximation, we can write our apparent position as a function of $\vec{r}$ and $\vec{u}$:

$$\vec{r}_{\rm app} = \vec{r} + \frac{1}{H}(\hat{\vec{r}} \cdot \vec{u})\hat{\vec{r}} \tag{226}$$
$$= \vec{r} + \frac{\mu u}{H}\hat{\vec{r}}, \tag{227}$$

where μ is the cosine of the angle between the velocity vector and the line of sight, i.e. $\mu = \hat{\vec{r}} \cdot \hat{\vec{k}}$ for a plane wave disturbance moving in the $\vec{k}$-direction.

- Now consider a plane-wave disturbance in the $\vec{k}$ direction, producing a displacement field $\vec{f}$ parallel to $\vec{k}$. This produces an apparent displacement

$$\vec{x} + f(\Omega)\mu x\hat{\vec{r}}, \tag{228}$$

and the component of this along the wave-vector is

$$x + f(\Omega)\mu^2 x. \tag{229}$$

In the framework of the Zel'dovich approximation, it is possible to show that the amplitude of the (apparent) density perturbation produced by this wave mode, δ_k, is proportional to the displacement, i.e.

$$\delta_{k,\rm true} \propto x \;\; \text{(the true displacement along the wave)} \tag{230}$$

and hence

$$\delta_{k,\rm app} \propto x + f(\Omega)\mu^2 x. \tag{231}$$

In other words, the apparent size of the density fluctuation is increased by a factor $1 + f(\Omega)\mu^2$, i.e. the fluctuation in redshift space, δ_s is related to the fluctuation in real space, δ_r, by:

$$\delta_s = \left[1 + f(\Omega)\mu^2\right]\delta_r. \tag{232}$$

- If the fluctuation in the number of galaxies (i.e. the light) is related to the fluctuation in the mass by some linear bias term, as argued above, then we can write the density perturbation in the light in redshift space as

$$\delta_s^{\rm light} = b\delta_s^{\rm mass} = \delta_s^{\rm mass} + (b-1)\delta_s^{\rm mass}. \tag{233}$$

The rearrangement emphasizes that there are two physically distinct contributions to $\delta_s^{\rm light}$ – one coming from dynamically generated density fluctuations, and which is sensitive to peculiar velocities, and a second which is due to the inherent clustering of the density peaks, and which is not affected by the peculiar velocities

- Therefore, in real space we have

$$\begin{aligned}
\delta_r^{\rm light} &= \delta_r^{\rm mass}\left[1+f(\Omega)\mu^2\right]+(b-1)\delta_r^{\rm mass} && (234)\\
&= \delta_r^{\rm light}\left[1+\frac{f(\Omega)\mu^2}{b}\right] && (235)\\
&= \delta_r^{\rm light}\left[1+\beta\mu^2\right]. && (236)
\end{aligned}$$

where in the last line, we have defined a new parameter $\beta \equiv f(\Omega)/b$.

- The foregoing analysis only applies in the linear regime. In the non-linear regime, the dominant effect is due to the virial motions of galaxies within clusters.

- This gives rise to an effect known as the **finger of God**

- Galaxies in the cluster that are moving away from us (e.g. falling in towards the cluster center of mass) have a large peculiar velocity away from us, and hence pick up a large peculiar redshift. On the other hand, galaxies at the back of the cluster that are falling in will have large peculiar velocities pointed towards us, and hence pick up blueshifts, thereby making them appear to be at lower redshifts.

- The effect is to distort the shape of the cluster in redshift space, elongating it along the line of sight.

- Cluster velocity dispersions are large, and so in the earliest redshift surveys (which surveyed only quite nearby volumes of the universe), this distortion lead to a pronounced elongation of each cluster, with all of the elongations apparently pointing at Earth! The first time this phenomenon was seen, it was humourously referred to as the "finger of God" pointing at us, and the name has since stuck.

- The influence of the "finger of God" effect on the redshift-space fluctuations can be approximately corrected for by damping each Fourier mode according to

$$\hat{\delta_k} \rightarrow \hat{\delta_k}\left(1+k^2\mu^2\sigma^2\right)^{-1/2}, \quad (237)$$

where σ is the velocity dispersion of the cluster.

- Combining this with the linear term, we find that the ratio between the power spectrum as measured in redshift space and its value in real space is simply

$$\frac{P_s}{P_r} = \frac{(1+\beta\mu^2)^2}{(1+k^2\mu^2\sigma^2)} \quad (238)$$